Verallgemeinerte CUSUM-Kontrollkarten und ihre Anwendung auf die Annahmestichprobenprüfung

Inaugural-Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Diplom-Mathematiker Ulrich Rendtel
aus Fürstenberg

Druckort: Berlin
Erscheinungsjahr: 1986

1986

Verallgemeinerte CUSUM-Kontrollkarten und ihre Anwendung auf die Annahmestichprobenprüfung

Inaugural-Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Diplom-Mathematiker Ulrich Rendtel
aus Fürstenberg

Druckort: Berlin
Erscheinungsjahr: 1986

Gedruckt mit Genehmigung des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

Fachbereichssprecher:	Professor Dr. W. Steden
Erstgutachter:	Professor Dr. H.-J. Lenz
Zweitgutachter:	Professor Dr. P.-Th. Wilrich

Tag der Disputation: 28. Juni 1985

Druck: Zentrale Universitäts-Druckerei der Freien Universität Berlin

Inhaltsverzeichnis

		Seite
1.	Einleitung und Überblick	1
2.	Motivation des CUSUM-Verfahrens und Herleitung von verallgemeinerten CUSUM-Kontrollkarten	14
2.1	Motivation des CUSUM-Verfahrens	14
2.2	Verallgemeinerte CUSUM-Karten	22
3.	Gütekriterien für den Entwurf von CUSUM-Karten	25
4.	Die Berechnung der Kenngrößen von verallgemeinerten CUSUM-Karten	28
4.1	Bestimmung der Übergangsdichte zwischen den transienten Zuständen des verallgemeinerten CUSUM-Schemas und die Berechnung der mehrstufigen Übergangsdichten M^n	28
4.2	Das asymptotische Verhalten von M^n und $P(N = n)$ im stationären Fall	36
4.3	Obere und untere Schranken für $P(N = n)$, $E(N)$, $VAR(N)$ und N_α	44
4.4	Lösung der Operatorgleichung für die ARL-Funktion	49
4.5	Die Berechnung der Kenngrößen bei instationärer Verteilungsannahme	54
5.	Das simultane Führen mehrerer CUSUM-Karten	60
5.1	Darstellung des Verfahrens als Markov-Kette mit mehrdimensionalem Zustandsraum	60
5.2	Berechnung der Lauflängenverteilung und der Gesamt-ARL im diskreten Fall	63
5.3	Berechnung der Gesamt-ARL im kontinuierlichen Fall	68
6.	CUSUM-Karten zur Kontrolle der Prozeßkurve	70
6.1	Modelle für die Stichprobenverteilung	70

		Seite
6.2	CUSUM-Kontrollkarten zum Aufdecken einer Prozeßverschlechterung	75
6.3	CUSUM-Karten zum Erkennen einer Prozeßverbesserung	93
7.	Adaptive Stichprobensysteme auf Grundlage von CUSUM-Karten	97
8.	Prüfstufensteuerung des MIL-STD-105 D mit Hilfe von verallgemeinerten CUSUM-Karten	116
8.1	Darstellung des Stichprobensystems MIL-STD-105 D und seiner Prüfstufensteuerung	116
8.2	Prüfstufensteuerung mit Hilfe von verallgemeinerten CUSUM-Karten	120
8.3	Berechnung der Kenngrößen von MIL-STD-105 D und seiner CUSUM-Versionen	125
8.4	Anpassen der Parameter der CUSUM-Versionen des MIL-STD-105 D	132
8.5	Vergleich der Lauflängenverteilung des MIL-STD-105 D und seiner CUSUM-Versionen	134
8.6	Vergleich von Prüfaufwand und Durchschlupf beim MIL-STD-105 D und der Binomial-CUSUM-Version	150
8.7	Abschließende Bemerkungen	160
	Literaturverzeichnis	162

Kapitel 1

Einleitung und Überblick

CUSUM-Karten wurden im Bereich der statistischen Qualitätskontrolle bisher bei der Prozeßkontrolle eingesetzt.

Bei der Prozeßkontrolle werden der laufenden Produktion Stichproben entnommen und anhand der vorliegenden Stichprobenergebnisse $X_t, X_{t-1}, \ldots, X_1$ wird entschieden, ob ein nicht direkt beobachtbarer Systemparameter ϑ_t zur Zeit t noch mit einem vorgegebenen Sollwert ϑ_S übereinstimmt. ϑ_t kann z. B. die mittlere Füllmenge sein, mit der eine Maschine zur Zeit t Flaschen abfüllt. Überschreitet ϑ_t einen vorgegebenen Wert ϑ_A, so soll das CUSUM-Verfahren diesen Wechsel von ϑ_S nach ϑ_A möglichst rasch aufdecken, was zu einem Justieren der Abfüllmaschine führt.

CUSUM-Kontrollkarten wurden 1954 von Page in der Form eines Intervallschemas eingeführt.

1.1) $$K_o = 0$$
$$K_t = \max \{ 0, K_{t-1} + X_t - k \} \quad t \geq 1$$

Eine Alarmmeldung erfolgt, falls:

1.2) $$K_t \geq h^K \quad t \geq 1$$

Hierbei sind k, $h^K > 0$ zu wählende Parameter, die die Eigenschaften der CUSUM-Karte bestimmen.

Besser interpretierbar ist jedoch eine Darstellung des CUSUM-Verfahrens über ein Likelihoodquotienten-Schema. Es sei hierzu $f(x|\vartheta_t)$ die Dichte von X_t an der Stelle x bei Vorliegen des Systemparameters ϑ_t. Das Likelihood-quotienten-Schema lautet dann:

$$(1.3) \qquad Q_o = 1$$
$$Q_t = \max \left\{ 1, Q_{t-1} \cdot \frac{f(X_t|\vartheta_A)}{f(X_t|\vartheta_s)} \right\} \quad t \geq 1$$

Eine Alarmmeldung erfolgt, falls:

$$(1.4) \qquad Q_t \geq h^Q \qquad t \geq 1$$

Falls die X_t aus einer 1-parametrigen Exponentialfamilie stammen, läßt sich das Q -Verfahren stets auf die Form (1.1) transformieren (vgl. Satz 2.2).

In der Form (1.3) ist das CUSUM-Schema als eine Folge von Sequentialquotiententests (SPRT's) mit der Untergrenze 1 und der Obergrenze h^Q interpretierbar (vgl. Khan (1978), (1984)). Jedesmal wenn ein einzelner SPRT die Untergrenze überschreitet, wird für den bisherigen Zeitabschnitt die Hypothese $\vartheta = \vartheta_s$ angenommen und ein neuer SPRT an der Untergrenze gestartet, bis schließlich der letzte SPRT die Obergrenze überschreitet, was zur Annahme der Hypothese $\vartheta = \vartheta_A$ führt.

In dieser Arbeit wird die folgende Verallgemeinerung des CUSUM-Schemas behandelt:

Im Sinne einer raschen Aufdeckung des Alarmzustandes erscheint es sinnvoll, immer dann informative Beobachtungen zu machen, wenn aufgrund des Likelihoodquotienten eine gewisse Information zugunsten des Alarmzustandes ϑ_A vorliegt. Umgekehrt wird man unnötigen Beobachtungsaufwand vermeiden wollen, wenn der Likelihoodquotient klein ist.

Ein verallgemeinertes CUSUM-Schema ist durch S Beobachtungsstufen gekennzeichnet. Hierbei sei $f(x|s_i, \sigma_t)$ $i = 1, \ldots, S$ die Dichte von X_t unter dem Systemparameter σ_t, wenn zur Zeit t die i-te Beobachtungsstufe vorliegt. Ein verallgemeinertes CUSUM-Schema sei dann zu vorgegebenen Werten $1 = h_1^Q < h_2^Q < \ldots < h_S^Q = h^Q$ definiert durch:

$$GQ_o = 1$$

$$(1.5) \qquad GQ_t = \max\{1, GQ_{t-1}\} \cdot \begin{cases} \frac{f(X_t|s_1, \sigma_A)}{f(X_t|s_1, \sigma_S)} & GQ_{t-1} < h_1^Q \\ \frac{f(X_t|s_2, \sigma_A)}{f(X_t|s_2, \sigma_S)} & h_1^Q \leq GQ_{t-1} < h_2^Q \\ \vdots & \vdots \\ \frac{f(X_t|s_S, \sigma_A)}{f(X_t|s_S, \sigma_S)} & h_{S-1}^Q \leq GQ_{t-1} < h_S^Q \end{cases}$$

Eine Alarmmeldung erfolgt, falls:

$$(1.6) \qquad GQ_t \geq h^Q \qquad t \geq 1$$

Für S = 1 gilt:

$$(1.7) \qquad GQ_t = \max\{1, Q_t\} \qquad t \geq 1$$

Folglich gibt GQ_t für S = 1 genau dann eine Alarmmeldung, wenn Q_t die Alarmgrenze h^Q überschreitet. Somit sind die beiden CUSUM-Verfahren für S = 1 äquivalent.

Die vorliegende Arbeit untersucht die Anwendung von einfachen und verallgemeinerten CUSUM-Karten im Bereich der Annahmestichprobenprüfung.

Hierbei wird die Situation behandelt, daß eine Folge von Warenlosen einer Warenausgangs- bzw. Wareneingangsprüfung unterzogen wird. Die Warenlose werden dabei jeweils mit einem einfachen (n, c) Prüfplan geprüft. In diesem

Fall sind die Beobachtungen X_t $(t \geq 1)$ die Anzahl der defekten Stücke in der Stichprobe aus dem t. Los.

Bei der Warenausgangsprüfung ist die Annahme-/Ablehnungsentscheidung häufig an die Prozeßkurve w_t gekoppelt, die bei der Produktion des t. Warenloses zugrunde liegt. Somit hängt die Korrektheit des verwendeten Prüfplans davon ab, ob die nicht direkt beobachtbare Prozeßkurve w_t noch mit dem ursprünglich angenommenen Sollwert w_s übereinstimmt. Ist dies aufgrund einer Alarmmeldung des CUSUM-Verfahrens nicht mehr der Fall, so muß entweder der Produktionsprozeß neu eingerichtet werden oder der Prüfplan muß neu angepaßt werden.

Eng hiermit verwandt ist das Problem, einen Adaptionsmechanismus zu finden, der einen bayesoptimalen Prüfplan (n, c) immer wieder an die aktuelle Prozeßkurve w_t angepaßt. In diesem Fall wird der Prüfplan (n, c) so gewählt, daß das Bayesrisiko bezüglich einer gegebenen Verlustfunktion unter Annahme einer Prozeßkurve w_s minimal wird (vgl. z. B. Hald (1981), Chap. 7). Ändert sich w_t, so muß auch (n, c) geändert werden.

Ein Adaptionsmechanismus beinhaltet die Lösung von 2 Problemen:
Zum einen muß entschieden werden, ob sich die Prozeßkurve geändert hat, und zum anderen muß die neue Prozeßkurve aus den zurückliegenden Stichprobenergebnissen $X_t, X_{t-1}, \ldots, X_1$ geschätzt werden.

Beschränkt man sich auf endlich viele Prozeßkurvenalternativen w_{A_i} $(i = 1, \ldots, N_A)$, dann können beide Probleme durch das simultane Führen von CUSUM-Karten zu den jeweiligen Alternativen $w_{A_i}(i = 1, \ldots, N_A)$ gelöst werden. Hierbei wird diejenige Alternative als neue Prozeßkurve gewählt, deren zugehörige CUSUM-Karte zum Alarmzeitpunkt den größten Wert hat. Wie gezeigt werden kann (vgl. Ka-

pitel 7) ist diese Schätzung eine ML-Schätzung unter den Alternativen w_{Ai} (i = 1, ..., N_A).

Bei der Wareneingangskontrolle ist die Situation gegeben, daß ein gewisser Prozentsatz p_S defekter Stücke in einem Los vom Abnehmer noch toleriert wird, während bei einem Prozentsatz $p_A > p_S$ die Warenannahme abgebrochen werden soll. Da dem Los nur Stichproben entnommen werden, wird $p_t = \sigma_t$ nicht direkt beobachtet.

Ziel eines Stichprobensystems in der Wareneingangskontrolle sollte es sein, den Prüfaufwand zu verringern, wenn die Stichprobenergebnisse der vorherigen Lose einen Produktionslevel $p_t \leq p_S$ andeuten. Zeigen umgekehrt die Stichprobenergebnisse eine Überschreitung von p_S an, so sollte der Stichprobenumfang vergrößert werden und damit die Stichprobe informativer gemacht werden.

Das Stichprobensystem MIL-STD-105 D (vgl. MILITARY STANDARD 105 D (1963)) verfügt über eine derartige Steuerung des Prüfaufwands. Der MIL-STD-105 D kennt drei Prüfstufen (reduzierte (R), normale (N) und verschärfte (T) Prüfung) mit drei Stichprobenumfängen $n_R < n_N \leq n_T$. Ein Nachteil des MIL-STD-105 D besteht in seiner komplizierten Prüfstufensteuerung (vgl. Abb. 8.1). So müssen die jeweils letzten 10 Stichprobenergebnisse bekannt sein, um über einen Übergang von normaler in die reduzierte Prüfung zu entscheiden.

Die Verallgemeinerung des CUSUM-Schemas (vgl. (1.5)) gestattet es, die Prüfstufensteuerung des MIL-STD-105 D auf das Führen einer einzigen (verallgemeinerten) CUSUM-Karte zu reduzieren. Im Hinblick auf das Verhältnis von Prüfaufwand (ASN) zu Durchschlupf (AOQ) ist diese CUSUM-Version des MIL-STD-105 D dem ursprünglichen Verfahren z. T. weit überlegen (vgl. Abschnitt 8.6).

Im folgenden soll ein kurzer Überblick über die einzelnen Teile dieser Arbeit gegeben werden.

Über die Art eines möglichen Wechsels des Systemparameters von $\mathcal{O}_S$ nach $\mathcal{O}_A$ existieren meistens nur vage Vorstellungen. Häufig wird man jedoch die Nichtprognostizierbarkeit des Sollzustandes akzeptieren, die besagt, daß aus Informationen darüber, daß sich das System im Sollzustand befindet, nicht geschlossen werden kann, daß sich das System auch weiterhin im Sollzustand befindet.

Wie in Kapitel 2 gezeigt wird, folgt aus der Nichtprognostizierbarkeit des Sollzustandes und dem Likelihoodprinzip, daß das Führen von CUSUM-Karten das natürliche Verfahren zur Kontrolle des Parameters $\mathcal{O}_t$ ist.

Der Rest von Kapitel 2 liefert die formale Definition von einfachen und verallgemeinerten CUSUM-Karten und zeigt einige ihrer fundamentalen Eigenschaften (Satz 2.1).

Kapitel 3 behandelt Gütekriterien, die die Wahl der freien Parameter des CUSUM-Verfahrens bestimmen. Üblicherweise fordert man, daß die erwartete Lauflänge N bis zu einem Alarm (Average Run Length = ARL) unter den Sollzustand groß und im Alarmzustand $\mathcal{O}_A$ klein sein soll.

Sind h^K und k in (1.1) und (1.2) frei wählbare Parameter, so fordert man, daß ihre Wahl das folgende Minimierungsproblem löst:

(1.8) minimiere $E_{\mathcal{O}_A}(N)$
k, h^K
unter $E_{\mathcal{O}_S}(N) = \text{const}$

Khan (1978) hat gezeigt, daß das Quotienten-Schema (1.3), bei dem k durch die Angabe von σ_A und σ_S über den Likelihoodquotienten bestimmt ist, eine Lösung des obigen Minimierungsproblems darstellt. Dies unterstützt noch einmal die Darstellung von CUSUM-Karten in der Form (1.3).

Der überwiegende Teil der Literatur über CUSUM-Karten hat sich mit der Angabe von Näherungen für die ARL (vgl. Johnson (1961), Johnson and Leone (1962 a + b), Reynolds (1975), Khan (1978)) oder mit der iterativen Bestimmung der ARL beschäftigt (vgl. z. B. v. Dobben de Bruyn (1968)). Diese ausschließliche Betrachtung der ARL wurde kritisiert zugunsten der Berücksichtigung der Quantile der Lauflängenverteilung (vgl. Barnard (1959), Bissell (1969).

Die Arbeiten von Ewan und Kemp (1960), Brook and Evans (1972), Zacks (1981), Woodall (1983), Waldmann (1984 a) und Waldmann (1984 b) haben sich mit der Berechnung der gesamten Verteilung von N befaßt.

Als Resultat der Ergebnisse von Kapitel 4 erhält man, daß die Verteilung von N auch bei verallgemeinerten CUSUM-Karten für ARL $\geq$ 50 gut durch eine geometrische Verteilung mit

$$\lambda = \frac{E(N)}{E(N)+1} \tag{1.9}$$

beschrieben wird. Die Lauflängenverteilung wird in diesem Fall durch einen einzigen Parameter, die ARL, völlig beschrieben. Diese Tatsache stützt das ARL-Kriterium.

Ein anderer Einwand gegen das ARL-Kriterium kann gegen die stationäre Verteilungsannahme bei der Berechnung von $E_{\sigma_A}(N)$ gemacht werden, die sich nicht mit der Modellannahme deckt, daß zu einem unbekannten Zeitpunkt ein Wechsel von σ_S nach σ_A stattgefunden hat. Wie in Kapitel 3 gezeigt wird,

stimmt jedoch $E_{\sigma_A}(N)$ ungefähr mit dem bedingten Erwartungswert der Lauflänge nach dem Sprung überein, gegeben daß ein Wechsel von σ_S nach σ_A stattgefunden hat.

Kapitel 4 beschäftigt sich mit der Berechnung der Verteilung von N für verallgemeinerte CUSUM-Karten. Es werden Schranken für $P(N = n)$ $(n \in \mathbb{N})$, $E(N)$, $VAR(N)$ und die α-Quantile N_α für den stationären und instationären Fall angegeben. Die Ergebnisse von Brook and Evans (1972), Woodall (1983), Waldmann (1983) und Waldmann (1984 a) lassen sich dabei auf verallgemeinerte CUSUM-Karten übertragen.

Die Darstellung in Kapitel 4 benutzt die Markov-Eigenschaft des mit ln transformierten verallgemeinerten CUSUM-Schemas:

$$(1.10)\qquad GCU_0 = 0 \qquad GCU_t = \max\{0, GCU_{t-1}\} + \begin{cases} Y_t(s_1) & GCU_{t-1} < h_1^{cu} = 0 \\ Y_t(s_2) & h_1^{cu} \le GCU_{t-1} < h_2^{cu} \\ \vdots & \vdots \\ Y_t(s_S) & h_{S-1}^{cu} \le GCU_{t-1} < h_S^{cu} \end{cases}$$

Es wird die Übergangsdichte M zwischen den transienten Zuständen des verallgemeinerten CUSUM-Schemas bestimmt. Die mehrstufigen Übergangsdichten M^n $(n \in \mathbb{N})$ lassen sich sukzessiv in gewohnter Weise bestimmen.

Das zentrale Hilfsmittel dieses Abschnitts ist eine verallgemeinerte Fassung des Theorems von Perron-Frobenius über das asymptotische Verhalten von M^n (vgl. Harris (63), p. 67). Man erhält unter gewissen Voraussetzungen an M:

$$(1.11)\qquad M^n(x,y) \xrightarrow[n\to\infty]{glm} \lambda^n \phi(x) \psi(y)$$

wobei ϕ die rechte und ψ die linke Eigenfunktion von M

zum maximalen Eigenwert λ sind. Die Sätze (4.2) und (4.3) geben hinreichende Bedingungen für die Gültigkeit von (1.11) an.

Als Konsequenz von (1.11) erhält man, daß die bedingte Verteilung von GCU_t ($t \geq 1$) auf den transienten Zuständen ergodisch ist und der Quotient zweier aufeinander folgender Terme der Dichte der Lauflänge N gegen λ kovergiert, folglich:

$$P(N = n + 1) \approx \lambda\, P(N = n) \tag{1.12}$$

Die Darstellung unter (1.12) motiviert zu einem Ansatz:

$$(c_n^-)^i\, P(N = n) \leq P(N = n+i) \leq (c_n^+)^i\, P(N = n) \quad i \in N \tag{1.13}$$

für obere und untere Schranken für die noch nicht berechneten Werte der Lauflängenverteilung. Über die in (1.13) gefundenen Abschätzungen lassen sich leicht Schranken für E(N), VAR(N) und die α-Quantile N_α gewinnen (Satz 4.8).

Will man lediglich die ARL berechnen, so ist es vorteilhaft, die Operatorgleichung

$$ARL(x) = 1 + \int M(x,y)\, ARL(y)\, \mu(dy) \tag{1.14}$$

zu lösen. Hierbei ist ARL(x) die erwartete Lauflänge, wenn das Verfahren bei $GCU_0 = x$ startet. Waldmann (1983) hat zur Lösung der Operatorgleichung (1.14) einen effektiven Extrapolationsansatz angegeben (vgl. Theorem 4.11).

Kapitel 5 behandelt den Fall, daß mehrere (einfache) CUSUM-Karten gleichzeitig geführt werden (vgl. Kapitel 7).

Das System der $(CU_t(1), \ldots, CU_t(N_A))$ $t \geq 0$

läßt sich als Markov-Kette mit mehrdimensionalem Zustandsraum auffassen.

Bei diskreten Zufallsgrößen X_t läßt sich eine Übergangsdichte M zwischen den transienten Zuständen sofort angeben. Die Ergebnisse des eindimensionalen Falls können dann übernommen werden, wenn (1.11) gilt. Hierfür werden hinreichende Bedingungen (Bed. 5.1 und 5.2) angegeben.

Bei kontinuierlichen Zufallsgrößen X_t ist der Ansatz über eine Übergangsdichte nicht ohne weiteres übertragbar. In diesem Fall läßt sich die ARL über eine Modifikation der Operatorgleichung (1.14) bestimmen.

Das 6. Kapitel behandelt den Fall, wie mit Hilfe einer (einfachen) CUSUM-Karte eine Veränderung der nicht beobachtbaren Prozeßkurve w_t erkannt werden kann.

Im Abschnitt 6.1 stellen wir ein Modell für den Zusammenhang zwischen der beobachteten Anzahl schlechter Stücke $X_t = x$ in der t. Stichprobe und w_t her. Man erhält:

$$(1.15) \qquad P(X_t = x) = f(x|n, w_t) = \int_0^1 b(x|n,p) w_t(p)\,dp \qquad x = 0, 1, \ldots, n$$

wobei b (x|n, p) die Dichte der Binomialverteilung an der Stelle x ist.

Im Abschnitt 6.2 untersuchen wir CUSUM-Karten, die eine Verschlechterung der Prozeßkurve erkennen. Als Prozeßkurve wird dabei die Dichte einer Betaverteilung gewählt. Für gewisse Parameteralternativen der Betaverteilung hat das CUSUM-Schema (1.3) eine besonders einfache und leicht interpretierbare Gestalt. Man erhält z. B. (vgl. (6.7):

(1.16) $Q_o = 1$

$$Q_t = \max\left\{1,\ Q_{t-1} \cdot \frac{E_s(Y|X_t = x_t)}{E_S(Y)}\right\}_{t \geq 1}$$

Hierbei ist $E_S(Y)$ die erwartete Anzahl schlechter Stücke im Restlos unter dem Sollzustand und $E_S(Y|X_t = x_t)$ der entsprechende bedingte Erwartungswert, wenn die Stichprobe X_t schlechte Stücke enthält.

Es wird das unterschiedliche Verhalten der ARL von 3 CUSUM-Karten untersucht, die zu Alternativen w_{Ai} (i = 1, 2, 3) gehören, die alle eine Verschlechterung des Produktionsprozesses beschreiben (vgl. Abb. 6.3.A1 - 6.3.A3).

Im Abschnitt 6.3 wird die Möglichkeit behandelt, CUSUM-Karten zu führen, die eine Verbesserung der Prozeßkurve erkennen.

In Kapitel 7 wird der in der Einleitung skizzierte Entwurf eines Adaptionsverfahrens für Bayes-Prüfpläne dargestellt. Das Verfahren wird diskutiert am Beispiel eines linearen Kostenmodells mit Prüf-, Ersatz- und Garantiekosten pro Einzelstück (vgl. Hald (1981), Chap. 7). Es wird die ARL des Gesamtsystems berechnet, die sich aus dem gleichzeitigen Führen von 3 CUSUM-Karten ergibt (vgl. Abb. 6.6 a + 6.6 b).

Es werden Gütekriterien für die Wahl der einzelnen Alarmgrenzen $h^{cu}(i)$ (i = 1, 2, 3) diskutiert.

Kapitel 8 zeigt, wie die Prüfstufensteuerung des MIL-STD-105 D durch verallgemeinerte CUSUM-Karten bewerkstelligt werden kann.

Legt man die von den Autoren des MIL-STD-105 D benutzte Binomial-Verteilung als Stichprobenverteilung zugrunde und setzt p_S = AQL und p_A = 3 AQL, so erhält man das folgende Schema:

$$GCU_o = 0$$

$$GCU_t = \max\{0,\ GCU_{t-1} + \begin{cases} Y_t(R) & GCU_{t-1} < 0 \\ Y_t(N) & 0 \leq GCU_{t-1} < h_T \\ Y_t(T) & h_T \leq GCU_{t-1} < h_D \end{cases}\} \tag{1.17}$$

$$\text{wobei: } Y_t(i) = \ln \frac{b(X_t \mid n_i,\ p_A)}{b(X_t \mid n_i,\ p_S)} \qquad i \in \{R,\ N,\ T\}$$

Ein Abbruch der Warenannahme erfolgt, wenn GCU_t den vorgegebenen Wert h_D überschreitet.

Im Abschnitt 8.2 werden noch 2 weitere CUSUM-Versionen des MIL-STD-105 D vorgestellt: Eine CUSUM-Prüfstufenregelung, die lediglich auf der Annahme-/Ablehnungsentscheidung der einzelnen Lose basiert und eine CUSUM-Prüfstufenregelung, die den Effekt von Prozeßvarianzen mitberücksichtigt.

Abschnitt 8.3 bestimmt die transienten Prüfzustände und die Übergangsdichte M beim MIL-STD-105 D (Satz 8.1) und beweist die Gültigkeit von (1.11) für M^n (Satz 8.2).

Die Wahl der Parameter h_T und h_D kann so erfolgen, daß die Lauflänge dieser CUSUM-Versionen des MIL-STD-105 D mit den ursprünglichen Verfahren für p = AQL übereinstimmt (vgl. Abschnitt 8.4).

Die Abschnitte 8.5 und 8.6 vergleichen die beiden Verfahren hinsichtlich ihres Verhaltens als Stoppregel

bschnitt 8.5) und hinsichtlich des Verhältnisses von
üfaufwand und Durchschlupf (Abschnitt 8.6).

s Resultat erhält man, daß die CUSUM-Versionen große
weichungen vom AQL-Wert wesentlich schneller erkennen
s der MIL-STD-105 D (vgl. Tabelle 8.2 und 8.6)

Berdem kommt die CUSUM-Version bei Durchschlupf-Werten
terhalb des AQL-Wertes mit einem erheblich geringeren
üfaufwand aus (ca. 60 % des Prüfaufwandes des MIL-STD-
5 D, vgl. Tabelle 8.8).

e obigen Resultate gelten noch schärfer, wenn man
att der Binomial-Verteilung die unter (1.15) angege-
ne gemischte Binomial-Verteilung als Stichprobenver-
ilung benutzt.

e in den Abschnitten 8.5 und 8.6 durchgeführte Unter-
chung über den Einfluß der Prozeßvarianz auf die Kenn-
ößen des MIL-STD-105 D ergänzt die bisher bekannten
beiten über den MIL-STD-105 D, die sein Verhalten für
n wenig realistischen Fall untersucht haben, daß die
gelieferten Lose alle denselben Losschlechtanteil p
ben (vgl. Brown and Rutemiller (1973), Koyama (1974),
yama (1978), Hald (1981), p. 288 - 298).

Kapitel 2

Motivation des CUSUM-Verfahrens und Herleitung von verallgemeinerten CUSUM-Kontrollkarten

Abschnitt 2.1: Motivation des CUSUM-Verfahrens

Wir nehmen an, daß wir an den Zeitpunkten t = 1, 2, ... eine Zufallsgröße X_t beobachten können. Die Beobachtung zum Zeitpunkt t sei abhängig von einem nicht direkt beobachtbaren Systemparameter $\vartheta_t \in \Theta$. Für $\vartheta_t = \vartheta_S$ sagen wir, daß sich das System zum Zeitpunkt t im Sollzustand befindet. Wir wollen annehmen, daß sich das System zu Beginn unserer Beobachtungen im Sollzustand ϑ_S befindet und zu einem unbekannten Zeitpunkt τ in einen Zustand ϑ_A überwechselt und in diesem Zustand verbleibt. Unser Ziel ist es, diesen Zustand ϑ_A, den wir Alarmzustand nennen, möglichst schnell zu entdecken.

Eine solche Situation ist z. B. bei der Annahmestichprobenprüfung gegeben: Aus einer Folge von Warenlosen wird jeweils eine Stichprobe entnommen. Hierbei ist X_t die Anzahl defekter Einzelstücke in der t. Stichprobe. Der Prozentsatz ϑ_t defekter Einzelstücke im t. Warenlos wird dabei nicht direkt beobachtet.

Bei der Wareneingangskontrolle ist häufig die Situation gegeben, daß ein gewisser Prozentsatz ϑ_S defekter Stücke

in einem Los vom Abnehmer noch toleriert wird, während bei einem Prozentsatz $\sigma_A > \sigma_S$ die Warenannahme abgebrochen werden soll.

Bei der Warenausgangskontrolle ist die Annahme-/Ablehnungsentscheidung über eine Lieferung häufig an die Prozeßkurve des Produktionsprozesses gekoppelt, aus dem die Einzelstücke stammen.

In diesem Fall ist σ_t die nicht direkt beobachtbare Prozeßkurve bei der Produktion des t. Loses. Ändert sich die Prozeßkurve vom angenommenen Sollwert σ_S, so muß auch der Prüfplan, mit dem die Warenlose kontrolliert werden, geändert werden.

Über die Art des Wechsels von σ_S nach σ_A existieren meistens nur vage Vorstellungen. Häufig wird jedoch folgender Sachverhalt akzeptiert:

(AI) "Aus dem Andauern des Sollzustandes kann kein Schluß auf das weitere Fortbestehen des Sollzustandes gezogen werden."

Das Akzeptieren von (AI) bedeutet insbesondere, daß keine Abhängigkeit der $(\sigma_t)_{t \geq 1}$ vorliegt. Würden etwa die σ_t einem autoregressiven Prozeß 1. Ordnung

$$\sigma_t = \varrho\, \sigma_{t-1} + \varepsilon_t \qquad 0 < \varrho < 1 \qquad \varepsilon_t \in N(0, \sigma^2)$$

genügen, so lassen sich aus Informationen über das Vorliegen des Sollzustandes zur Zeit t Prognosen für σ_{t+1} gewinnen. Bei Campling (68), Cox (70), Both and Smith (76), Reimann (84) und Reetz (84) werden Situationen betrachtet, wo die $(\sigma_t)_{t \geq 1}$ eine Markov-Kette bilden.

Vom informationstheoretischen Aspekt gesehen ist die unter (AI) beschriebene Situation ungünstiger als bei Abhängigkeit der $(\sigma_t)_{t \geq 1}$.

Die Situation wird dadurch etwas günstiger, daß wir uns auf die Annahme beschränken, daß σ_t zu einem unbekannten Zeitpunkt τ von σ_S auf σ_A springt und dort verbleibt:

(AII) $$\sigma_t = \begin{cases} \sigma_S & t \leq \tau \\ \sigma_A & t > \tau \end{cases} \qquad \tau \text{ unbekannt}$$

Wir wollen weiterhin annehmen, daß die Beobachtungsprozesse zu den einzelnen Zeitpunkten t (t = 1, 2, ...) unabhängig voneinander sind:

(AIII)
Die X_t (t = 1, 2, ...) sind unabhängig voneinander.

Es sei $f_{X_t}(x|\sigma_t) = f(x|\sigma_t)$ $(\sigma_t \in \mathfrak{S})$ die Dichte von X_t bezüglich eines geeigneten Maßes. Falls wir bis zu einem Zeitpunkt t die Beobachtungen $X_1 = x_1, \ldots, X_t = x_t$ gemacht haben, erhalten wir für den Likelihoodquotienten q_t:

$$q_t = \prod_{t'=1}^{t} \frac{f(x_{t'}|\sigma_A)}{f(x_{t'}|\sigma_S)}$$

Das Likelihoodprinzip besagt: Falls $q_t < 1$ gilt, so spricht dies eher für das Vorliegen von σ_S, während für große Werte von q_t der Alarmzustand σ_A sehr wahrscheinlich eingetreten ist.

Gemäß (AI) müssen wir im ersten Fall q_t durch den "neutralen" Wert 1 ersetzen. Wir erhalten damit:

Definition 2.1:

Eine CUSUM-Kontrollkarte ist für gegebene Parameter ϑ_S, ϑ_A und h^Q definiert durch:

$$Q_0 = 1$$

$$Q_t = \max\left\{1, Q_{t-1} \cdot \frac{f(X_t \mid \vartheta_A)}{f(X_t \mid \vartheta_S)}\right\} \qquad t = 1,2,\dots \tag{2.1}$$

$$\text{Alarm falls: } Q_t \geq h^Q \qquad t = 1, 2, \dots \tag{2.2}$$

□

Wie wir gesehen haben, ist das CUSUM-Verfahren eine unmittelbare Folgerung aus dem Likelihood-Prinzip und [illegible].

Als Lauflänge N einer CUSUM-Kontrollkarte definieren wir:

$$N := \inf\{t \mid Q_t \geq h^Q,\ t \in \mathbb{N}\}$$

Als Folgerung aus der Definition der Q_t erhalten wir:

Satz 2.1:

$$Q_t = \max\left\{1, \max_{1 \leq \ell \leq t} \prod_{t'=\ell}^{t} \frac{f(X_{t'} \mid \vartheta_A)}{f(X_{t'} \mid \vartheta_S)}\right\} \qquad t = 1,2,\dots \tag{2.3}$$

Beweis:

Durch Induktion: (2.3) ist für t = 1 erfüllt. Unter Annahme, daß Q_t (2.3) erfüllt, erhalten wir:

$$Q_{t+1} = \max\left\{1, \frac{f(x_{t+1}|\vartheta_A)}{f(x_{t+1}|\vartheta_S)} \max\left\{1, \max_{1\le l\le t} \prod_{i=l}^{t} \frac{f(x_i|\vartheta_A)}{f(x_i|\vartheta_S)}\right\}\right\}$$

$$= \max\left\{1, \frac{f(x_{t+1}|\vartheta_A)}{f(x_{t+1}|\vartheta_S)} \cdot \max_{1\le l\le t} \prod_{i=l}^{t} \frac{f(x_i|\vartheta_A)}{f(x_i|\vartheta_S)}\right\}$$

$$= \max\left\{1, \max_{1\le l\le t+1} \prod_{i=l}^{t+1} \frac{f(x_i|\vartheta_A)}{f(x_i|\vartheta_S)}\right\}$$

□

Das CUSUM-Verfahren wählt also zu jedem Zeitpunkt t gerade so viele der zurückliegenden Beobachtungen x_t, x_{t-1}, x_{t-2}, ... aus, daß der Likelihoodquotient maximal wird. Es ist eine dynamische Steuerung der Anzahl der Beobachtungen, die in eine Entscheidung über eine Alarmmeldung einfließen.

Es ist klar, daß jeder Kontrollmechanismus neuere Beobachtungen stärker berücksichtigen muß als ältere. Bei den übrigen mit dem CUSUM-Verfahren konkurrierenden Kontrollkarten ist diese Regelung jedoch starr: Bei den Shewhart-Kontrollkarten wird der Likelihoodquotient aus dem jeweils letzten Stichprobenergebnis gebildet. Bei den sogenannten MOSUM-Karten gehen die jeweils letzten m Beobachtungen in die Berechnung des Likelihoodquotienten ein, wobei m vorgegeben ist, vgl. z. B. Roberts (66), Lai (74), Bauer und Hackl (78).

Das unter Definition 2.1 beschriebene CUSUM-Verfahren ist identisch mit der Durchführung einer Folge von Sequentialquotiententests (SPRT's) mit der Untergrenze 1 und der Obergrenze h^Q. Jedesmal, wenn ein SPRT die Untergrenze überschreitet, wird ein neuer SPRT bei 1 gestartet, bis ein SPRT schließlich die Obergrenze über-

schreitet. Die Eigenschaften des CUSUM-Verfahrens lassen sich damit über die Eigenschaften des mitbenutzten SPRT herleiten, vgl. z. B. Johnson (1961), Johnson and Leone (1962), Khan (1978) und Khan (1984).

Durch $CU_t = \ln(Q_t)$ $t = 0, 1, 2, \ldots$ läßt sich das Quotientenschema auf ein Summationsschema transformieren. Man erhält:

(2.4) $$CU_o = 0$$
$$CU_t = \max \{ 0, CU_{t-1} + Y_t \} \quad t = 1, 2, \ldots$$

wobei $Y_t := \ln \frac{f(X_t|\vartheta_t)}{f(X_t|\vartheta_S)}$, und

(2.5) Alarm, falls: $CU_t \geq h^{cu}$ $t \geq 1$

wobei $h^{cu} := \ln(h^Q)$.

Für den Fall, daß die Beobachtungen X_t aus einer 1-parametrigen Exponentialfamilie stammen, d. h. $f(x|\vartheta)$ ist von der Form

(2.6) $$f(x|\vartheta) = g(x)h(\vartheta) \exp(x\vartheta)$$

läßt sich das Quotientenschema auf das klassische (k, h)-Schema transformieren (vgl. z. B. de Bruyn (1968)).

Über die Transformation

(2.7) $$\frac{1}{\vartheta_A - \vartheta_S} \ln \frac{f(x|\vartheta_A)}{f(x|\vartheta_S)} = X + \frac{\ln \frac{h(\vartheta_A)}{h(\vartheta_S)}}{\vartheta_A - \vartheta_S}$$

erhält man die transformierten Schemata:

$$K_o^+ = K_o^- = 0$$

(2.8) $$K_t^+ = \max \{ 0, K_{t-1}^+ + X_t - k \} \quad t = 1, 2, \ldots \quad \sigma_A > \sigma_S$$

(2.9) $$K_t^- = \min \{ 0, K_{t-1}^- + X_t + k \} \quad t = 1, 2, \ldots \quad \sigma_A < \sigma_S$$

Es wird Alarm gegeben, falls:

(2.10) $$\begin{aligned} K_t^+ &\geq h^K \quad t \geq 1 \quad \sigma_A > \sigma_S \\ K_t^- &\leq -h^K \quad t \geq 1 \quad \sigma_A < \sigma_S \end{aligned}$$

Satz 2.2:

Für $\sigma_A > \sigma_S$ bzw. $\sigma_A < \sigma_S$ sind das Q_t-Verfahren und die K_t^+ - bzw. K_t^- - Kontrollkarte äquivalent, falls:

(2.11) $$k = \frac{-\ln \frac{h(\sigma_A)}{h(\sigma_S)}}{|\sigma_A - \sigma_S|}$$

(2.12) $$h^K = \frac{\ln (h^Q)}{|\sigma_A - \sigma_S|}$$

Beweis:
Setze $m(g) := \frac{\ln(g)}{\sigma_A - \sigma_S}$. Für $\sigma_A > \sigma_S$ ist $m(g)$ streng isoton in g und es folgt aus (2.7):

$$m\left(\frac{f(x|\sigma_A)}{f(x|\sigma_S)}\right) = x - k$$

Für $\sigma_A > \sigma_S$ folgt dann die Äquivalenz der Q_t- und K_t^+ - Karte aus

$$m(Q_t) = K_t^+ \qquad t \geq 0$$

was man sofort über vollständige Induktion zeigt.

Für $\sigma_A < \sigma_S$ ist $-m(g)$ streng isoton und es folgt aus (2.7):

$$-m\left(\frac{f(x|\sigma_A)}{f(x|\sigma_S)}\right) = -(x + k)$$

Die Äquivalenz der Q_t- und K_t^--Karte folgt dann aus:

$$m(Q_t) = K_t^- \qquad t \geq 1$$

und aus:

$$-\min(-a, -b) = \max(a, b) \qquad a, b \in \mathbb{R}$$

□

Neben dem Intervall-Schema existiert noch eine graphische Version des CUSUM-Verfahrens, die sog. V-Masken-Technik (vgl. Barnard (1959)), die mit dem CUSUM-Verfahren äquivalent ist.

Die bei Bissell (1979) und Hackl (1981) genannten Verallgemeinerungen der V-Masken, sog. semiparabolische Masken, sind nicht mehr mit dem CUSUM-Verfahren äquivalent.

Abschnitt 2.2: Verallgemeinerte CUSUM-Karten

Bisher haben wir den Fall betrachtet, daß die Verteilung von X_t ($t \geq 1$) nur von σ_t abhängt. Im Fall der Annahmestichprobenprüfung bedeutet dies, daß der Stichprobenumfang n konstant ist.

Im Sinne einer raschen Aufdeckung des Alarmzustandes erscheint es jedoch sinnvoll, immer dann informativere Beobachtungen zu machen, wenn aufgrund des Likelihoodquotienten eine gewisse Information zugunsten des Alarmzustandes σ_A vorliegt. Im Fall der Annahmestichprobenprüfung heißt dies, den Stichprobenumfang n zu vergrößern, je größer Q_t bzw. CU_t ist.

Umgekehrt wird man unnötigen Prüfaufwand vermeiden wollen, wenn Q_t bzw. CU_t klein ist.

Ein verallgemeinertes CUSUM-Schema ist durch S Stufen gekennzeichnet. Hierbei sei $f(x|s_i, \sigma_t)$ $i = 1, \ldots, S$ die Dichte von X_t unter dem Systemparameter σ_t, wenn zur Zeit t die i. Beobachtungsstufe vorliegt.

Definition 2.2:

Ein verallgemeinertes CUSUM-Schema ist zu vorgegebenen Werten $1 = h_1^Q < h_2^Q < \ldots < h_S^Q = h^Q$ definiert durch:

$$GQ_0 = 1$$

$$(2.13)\quad GQ_t = \max\{1, GQ_{t-1}\} \cdot \begin{cases} \frac{f(X_t|\sigma_1, \sigma_A)}{f(X_t|\sigma_1, \sigma_S)} & GQ_{t-1} < h_1^Q \\ \frac{f(X_t|\sigma_2, \sigma_A)}{f(X_t|\sigma_2, \sigma_S)} & h_1^Q \leq GQ_{t-1} < h_2^Q \\ \vdots & \vdots \\ \frac{f(X_t|\sigma_S, \sigma_A)}{f(X_t|\sigma_S, \sigma_S)} & h_{S-1}^Q \leq GQ_{t-1} < h_S^Q \end{cases}$$

$$t \geq 1$$

(2.14) Alarm falls: $GQ_t \geq h^Q$, $t \geq 1$ □

Bemerkung 2.1:

Das CUSUM-Schema Q_t aus Definition 2.1 ist mit dem verallgemeinerten CUSUM-Schema für S = 1 äquivalent. In diesem Fall lautet (2.13):

(2.15) $$GQ_t = \max\{1, GQ_{t-1}\} \cdot \frac{f(X_t|\vartheta_1)}{f(X_t|\vartheta_0)} \qquad t \geq 1$$

Zwar kann $(GQ_t)_{t\geq 1}$ im Gegensatz zu $(Q_t)_{t\geq 1}$ auch Werte kleiner als 1 annehmen. Wegen

(2.16) $$Q_t = \max\{1, GQ_t\}$$

überschreitet GQ_t genau dann die h^Q-Alarmgrenze, wenn Q_t diese Grenze überschreitet. Folglich stimmt die Lauflänge N für beide Verfahren für S = 1 überein. □

Durch $GCU_t := \ln(GQ_t)$ t = 0, 1, 2, ... läßt sich das Quotientenschema wieder auf ein Summationsschema transformieren. Man erhält:

$$GCU_0 = 0$$

(2.17) $$GCU_t = \max\{0, GCU_{t-1}\} + \begin{cases} Y_t(\rho_1) & GCU_{t-1} < h_1^{cu} \\ Y_t(\rho_2) & h_1^{cu} \leq GCU_{t-1} < h_2^{cu} \\ \vdots & \vdots \\ Y_t(\rho_S) & h_{S-1}^{cu} \leq GCU_{t-1} < h_S^{cu} \end{cases}$$

$$t \geq 1$$

wobei $$Y_t(\rho_i) := \ln \frac{f(X_t|\rho_i, \vartheta_1)}{f(X_t|\rho_i, \vartheta_0)} \qquad i = 1, \ldots, S$$

$$h_i^{cu} := \ln(h_i^Q) \qquad i = 1, \ldots, S$$

Beispiel 2.1:

Die beobachteten X_t $(t \geq 1)$ seien binomial-verteilt. Der erwartete Prozentsatz p defekter Stücke betrage im Sollzustand p_S und im Alarmzustand $p_A > p_S$ Prozent. Eine Beobachtungsstufe s_i ist dann durch einen Stichprobenumfang $n(i)$ $i = 1, \ldots, S$ gegeben.

Im allgemeinen wird man $n(1) \leq n(2) \leq \ldots \leq n(S)$ wählen. Die Wahl des Stichprobenumfangs n_t für die t. Beobachtung wird gemäß (2.17) aufgrund des vorhergehenden Wertes GCU_{t-1} bestimmt.

Durch die Transformation

$$GK_t := \frac{1}{\ln \frac{p_A(1-p_S)}{p_S(1-p_A)}} GCU_t \qquad t \geq 0 \tag{2.18}$$

läßt sich das verallgemeinerte CUSUM-Schema auf eine einfache Form bringen:

$$GK_t = \max\{0, GK_{t-1}\} + \begin{cases} X_t - k(n(1)) & GK_{t-1} < h_1^K \\ X_t - k(n(2)) & h_1^K \leq GK_{t-1} < h_2^K \\ \vdots & \vdots \\ X_t - k(n(S)) & h_{S-1}^K \leq GK_{t-1} < h_S^K \end{cases} \qquad t \geq 1 \tag{2.19}$$

Hierbei ist:

$$k(n) := \frac{-\ln \frac{1-p_A}{1-p_S}}{\ln \frac{p_A(1-p_S)}{p_S(1-p_A)}} \cdot n \tag{2.20}$$

$$h_i^K := \frac{h_i^{CU}}{\ln \frac{p_A(1-p_S)}{p_S(1-p_A)}} \qquad i = 1, \ldots, S \tag{2.21}$$

□

Kapitel 3

Gütekriterien für den Entwurf von CUSUM-Karten

Bei der Wahl der Parameter der CUSUM-Karte fordert man üblicherweise, daß die mittlere Laufzeit (Average Run Length = ARL) $E_{\sigma}(N)$ im Sollzustand $\sigma = \sigma_S$ groß und im Alarmzustand $\sigma = \sigma_A$ klein ist.

Bei dem (k, h)-Intervall-Schema (vgl. (2.9), (2.10)) werden k und h als frei wählbare Parameter behandelt.

In der Interpretation durch den Likelihoodansatz ist k jedoch durch die Wahl einer Alternative σ_A festgelegt, wobei (vgl. Satz 2.2)

(3.1) $$k = \frac{-\ln \frac{h(\sigma_A)}{h(\sigma_S)}}{|\sigma_A - \sigma_S|}$$

Falls k und h als frei wählbar angesehen werden, fordert man, daß ihre Wahl das folgende Minimierungsproblem löst:

(3.2) minimiere $\underset{k,\,h}{} E_{\sigma_A}(N)$

unter $E_{\sigma_S}(N) = \text{const.}$

Khan (1978) hat gezeigt, daß (3.1) die optimale Wahl von k im Sinne des obigen Minimierungsproblems ist.

Umgekehrt bedeutet jede Wahl von k eine implizite, über (3.1) definierte Wahl einer Alternative σ_A.

Für verallgemeinerte CUSUM-Karten gilt dieser Zusammenhang jedoch nicht immer, vgl. Beispiel 2.1 .

Gegen das ARL-Kriterium sind verschiedene Einwände gemacht worden. Barnard (1959) hat darauf hingewiesen, daß die ausschließliche Betrachtung des Mittelwertes von N nicht sinnvoll ist, da die Folgekosten für das Nichterkennen des Alarmzustandes häufig nicht proportional zu N sind. Es sei daher sinnvoller, die α-Prozentpunkte N_α der Verteilung von N zu betrachten. Auch die Nichtbeachtung der hohen Varianz von N sei nicht praktikabel.

In dieser Arbeit wird jedoch gezeigt, daß die Verteilung von N für $E(N) \geq 50$ gut durch eine geometrische Verteilung approximiert werden kann. In diesem Fall gilt:

$$VAR(N) \approx E^2(N) \tag{3.3}$$

$$N_\alpha \approx \frac{\ln(1-\alpha)}{\ln \frac{E(N)}{E(N)+1}} - 1 \tag{3.4}$$

Damit bedeutet aber die Kontrolle von E(N) die gleichzeitige Kontrolle von VAR(N) und N_α, wobei VAR(N) und N_α isoton von E(N) abhängen.

Ein anderer Einwand kann gegen die stationäre Verteilungsannahme bei der Formulierung des ARL-Kriteriums gemacht werden.

Wir betrachten hierzu den einfachsten instationären Fall, wo der Parameter σ_t zu einem unbekannten Zeitpunkt τ von σ_S nach σ_A wechselt.

Es sei K_τ die Laufzeit der verallgemeinerten CUSUM-Karte nach dem Wechsel von σ_S zu σ_A:

$$K_\tau := \begin{cases} 0 & N \leq \tau \\ k & N = \tau + k \qquad k \in \mathbb{N} \end{cases}$$

Dann ist plausibel zu fordern:

(3.5) $\quad E(K_\tau | N > \tau) = \text{klein} \qquad$ für alle τ

Das obige Kriterium sieht zunächst wenig praktikabel aus, da $E(K_\tau | N > \tau)$ vom unbekannten τ abhängt. Wie jedoch gezeigt werden kann, konvergiert $E(K_\tau | N > \tau)$ sehr rasch und ist für $\tau > 30$ fast konstant (vgl. Satz 4.14). Die Konvergenz von $E(K_\tau | N > \tau)$ hat seine Ursache darin, daß die bedingte Verteilung auf den Zuständen, die die CUSUM-Karte vor einem Stopp annehmen kann, konvergiert.

$E_{\sigma_A}(N)$ kann dann aufgefaßt werden als ein bedingter Erwartungswert, wobei die bedingte Verteilung auf den transienten Zuständen zum Zeitpunkt τ durch eine Verteilung ersetzt wird, die auf dem Startzustand 0 konzentriert ist.

Wie die numerischen Beispiele zeigen (vgl. Tab. 8.6), sind die Unterschiede zwischen $E_{\sigma_A}(N)$ und $E(K_\tau | N > \tau)$ ($\tau > 30$) klein, so daß das ARL-Kriterium auch unter diesem Aspekt sinnvoll bleibt.

Kapitel 4

Die Berechnung der Kenngrößen von verallgemeinerten CUSUM-Karten

Abschnitt 4.1:
Bestimmung der Übergangsdichte zwischen den transienten Zuständen des CUSUM-Schemas und die Berechnung der mehrstufigen Übergangsdichten M^n

Ein verallgemeinertes CUSUM-Schema mit

$$GCU_0 = 0$$

$$GCU_t = \max\{0, GCU_{t-1}\} + \begin{cases} Y_t(s_1) & GCU_{t-1} < h_1^{cu} \\ Y_t(s_2) & h_1^{cu} \leq GCU_{t-1} < h_2^{cu} \\ \vdots & \vdots \\ Y_t(s_S) & h_{S-1}^{cu} \leq GCU_{t-1} < h_S^{cu} \end{cases}$$

$$t \geq 1$$

bildet eine Markov-Kette.

Das CUSUM-Schema befindet sich zum Zeitpunkt t im Zustand E_z, falls $GCU_t = z$ gilt. Die Zustände mit $z \geq h^{cu} = h_S^{cu}$ werden zu einem absorbierenden Zustand E_H zusammengefaßt. Die Zustände E_z mit $z < 0$ werden zu einem Zustand E_{-1} zusammengefaßt. Der Zustand E_{-1} und die Zustände E_z mit $z \in [0, h^{cu})$ bilden die transienten Zustände der Markov-Kette. Dies sind alle Zustände, die das CUSUM-Schema vor Überschreiten der Alarmgrenze erreichen kann.

Läßt sich der Wertebereich aller $Y_t(s_i)$ $i = 1, \ldots, S$ durch Multiplikation mit einer Konstanten als Teilmenge

der ganzen Zahlen darstellen, so ist der Zustandsraum der Markov-Kette endlich. Die transienten Zustände E_{z_i} ($i = 1, 0, \ldots, H-1$) sind dann repräsentiert durch:

$$z_{-1} = -1 < z_0 = 0 < z_1 < \ldots < z_{H-1} < h^{cu}$$

Hierbei sind die z_i ($i = 1, \ldots, H-1$) diejenigen positiven Werte, die das CUSUM-Schema vor einem Stopp erreichen kann.

Dieser endliche Fall ist z. B. gegeben, wenn sich das CUSUM-Schema auf die Form

$$(4.2)\quad GK_t = \max\{0, GK_{t-1}\} + \begin{cases} X_t - \frac{n_1}{m_1} & GK_{t-1} < h_1^K \\ X_t - \frac{n_2}{m_2} & h_1^K \le GK_{t-1} < h_2^K \\ \vdots & \vdots \\ X_t - \frac{n_S}{m_S} & h_{S-1}^K \le GK_{t-1} < h_S^K \end{cases}$$

$$n_i, m_i \in \mathbb{N} \quad , \quad i = 1, \ldots, S$$

transformieren läßt und die X_t nur ganzzahlige Werte annehmen (vgl. Beispiel 2.1).

Im kontinuierlichen Fall läßt sich die Menge der transienten Zustände durch das Intervall $[0, h^{cu})$ und den Punkt -1 parametrisieren.

Wir wollen nun die Übergangsdichte M_t zum Zeitpunkt t zwischen den transienten Zuständen des CUSUM-Schemas bestimmen. Für jedes $x \in [0, h^{cu})$ und jede meßbare Teilmenge A von $[0, h^{cu})$ muß M_t der folgenden Gleichung genügen:

$$(4.3)\quad P(GCU_t \in A \mid GCU_{t-1} = x) = \int_A M_t(x, z)\, \nu(dz)$$

wobei ν ein geeignetes Maß auf $[o, h^{cu})$ ist.

Für $x = -1$ muß $M_t(-1, z)$ die Gleichung

$$(4.4) \qquad P(GCU_t \in A | GCU_{t-1} < 0) = \int_A M(-1, z)\, \nu(dz)$$

erfüllen. Für $M_t(-1, -1)$ muß gelten:

$$(4.5) \qquad P(GCU_t < 0 | GCU_{t-1} < 0) = M_t(-1, -1)$$

Schließlich muß für den Übergang von E_x mit $x \in [o, h^{cu})$ nach E_{-1} gelten:

$$(4.6) \qquad P(GCU_t < 0 | GCU_{t-1} = x) = M_t(x, -1)$$

Ist die Menge der transienten Zustände endlich, so kann in (4.3) ν als das auf $\{z_o, z_1, \ldots, z_{H-1}\}$ konzentrierte Zählmaß gewählt werden. Im kontinuierlichen Fall ist ν das auf $[0, h^{cu})$ eingeschränkte Lebesgue-Maß λ.

Es sei $g_t(y|s_i)$ die Dichte der Zuwachsfunktion

$$Y_t(\rho_i) = \ln \frac{f(X_t | \rho_i, \vartheta_1)}{f(X_t | \rho_i, \vartheta_S)} \qquad i = 1, \ldots, S$$

zum Zeitpunkt t.

Aus (4.1) folgt unmittelbar:

$$(4.7) \quad M_t(z, -1) = \begin{cases} P(Y_t(\rho_1) < 0) & z = -1 \\ P(Y_t(\rho_2) < -z) & 0 \le z < h_2^{cu} \\ \vdots & \vdots \\ P(Y_t(\rho_S) < -z) & h_{S-1}^{cu} \le z < h_S^{cu} \end{cases}$$

Für $x \in [0, h^{cu})$ erhält man:

$$(4.8)\quad M_t(z,x) = \begin{cases} g_t(x \mid \rho_1) & z = -1 \\ g_t(x-z \mid \rho_2) & 0 \le z < h_2^{cu} \\ \vdots & \vdots \\ g_t(x-z \mid \rho_S) & h_{S-1}^{cu} \le z < h_S^{cu} \end{cases}$$

Bemerkung 4.1:

Im Fall einer einfachen CUSUM-Karte mit $CU_0 = 0$ und

$$(4.9)\quad CU_t = \max\{0,\ CU_{t-1} + Y_t\} \qquad t \geq 1$$

entfällt der transiente Zustand E_{-1}. Für jedes $x \in [0,\ h^{cu})$ und jede meßbare Teilmenge A von $[0,\ h^{cu})$ genügt die Übergangsdichte $\widetilde{M}_t$ des CUSUM-Schemas unter (4.9) der Gleichung

$$(4.10)\quad P(CU_t \in A \mid CU_{t-1} = x) = \int_A \widetilde{M}_t(x,\ z)\ \widetilde{\nu}(dz)$$

Hierbei ist $\widetilde{\nu}$ entweder durch das auf $\{z_0,\ z_1,\ \ldots, z_{H-1}\}$ konzentrierte Zählmaß oder die Summe des Zählmaßes auf $z_0 = 0$ und des auf $[0,\ h^{cu})$ eingeschränkten Lebesque-Maßes λ gegeben.

Aus (4.9) folgt unmittelbar:

$$(4.11)\quad \widetilde{M}_t(x,\ z) = \begin{cases} P(Y_t \le -x) & z = 0 \\ g_t(z-x) & 0 < z < h^{cu} \end{cases}$$

□

Für die meisten Anwendungen ist es nötig, sich auf endlich viele Zustände zu beschränken. Es sei hierzu das Intervall $[0,\ h^{cu})$ in endlich viele Teilintervalle der Länge Δ zerlegt mit

$$I_z := [z,\ z+\Delta) \qquad z = 0, \Delta, \ldots, (h^{cu}-1)\Delta$$

Das CUSUM-Schema befindet sich dann zum Zeitpunkt t im Zustand E_z, falls $GCU_t \in I_z$ gilt. Dieses Verfahren entspricht dem Runden der Werte von GCU_t.

In diesem Fall läßt sich die Übergangsdichte M_t wie folgt approximieren (vgl. Brook and Evans (1972)):

Für z, x $\in \{0, \Delta, \ldots, (h^{cu}-1)\Delta\}$ wählt man:

$$P(GCU_t \in I_z \mid GCU_{t-1} \in I_x) = P(GCU_t \in I_z \mid GCU_{t-1} \in [x, x+\Delta)) \approx P(GCU_t \in I_z \mid GCU_{t-1} = x + \tfrac{\Delta}{2}) \tag{4.12}$$

Dieser Ansatz liefert:

$$M_t(x,z) \approx \begin{cases} P(z \le Y_t(\sigma_2) + \frac{\Delta}{2} + x < z + \Delta) & 0 \le x + \frac{\Delta}{2} < h_2^{cu} \\ \vdots & \vdots \\ P(z \le Y_t(\sigma_S) + \frac{\Delta}{2} + x < z + \Delta) & h_{S-1}^{cu} \le x + \frac{\Delta}{2} < h_S^{cu} \end{cases} \tag{4.13}$$

$$= \begin{cases} P(Y_t(\sigma_2) + \frac{\Delta}{2} \in I_{z-x}) & 0 \le x + \frac{\Delta}{2} < h_2^{cu} \\ \vdots & \vdots \\ P(Y_t(\sigma_S) + \frac{\Delta}{2} \in I_{z-x}) & h_{S-1}^{cu} \le x + \frac{\Delta}{2} < h_S^{cu} \end{cases}$$

Für den Übergang von einem Intervall I_x in den Zustand E_{-1} erhält man:

$$M_t(x,-1) \approx \begin{cases} P(Y_t(\sigma_2) < -(x + \frac{\Delta}{2})) & 0 \le x + \frac{\Delta}{2} < h_2^{cu} \\ \vdots & \vdots \\ P(Y_t(\sigma_S) < -(x + \frac{\Delta}{2})) & h_{S-1}^{cu} \le x + \frac{\Delta}{2} < h_S^{cu} \end{cases} \tag{4.14}$$

Schließlich erhält man für den Übergang vom Zustand E_{-1} in ein Intervall I_z:

$$M_t(-1, z) = P(Y_t(s_1) \in I_z) \qquad z = 0, \Delta, \ldots, (h^{cu} - 1)\Delta \tag{4.15}$$

Völlig analog läßt sich eine Diskretisierung des Zustandsraums für das CU_t-Schema durchführen.

Der Einfluß von $\Delta = \frac{h^{cu}}{N_{Gitter}}$ auf die

Berechnung der ARL sei an folgendem Beispiel verdeutlicht:

Beispiel 4.1:

Die Dichte von X_t sei durch eine Beta-Binomial-Verteilung (= Polya-Verteilung) gegeben mit n = 125, vgl. (6.5). Im Sollzustand habe die mischende Betaverteilung den Erwartungswert $\bar{p}_S$ = 0.9 % und die Streuung σ_S = 0.47 %. Im Alarmzustand gelte $\bar{p}_A$ = 1.3 % und σ_A = 1.0 %.

Tabelle 4.1 zeigt die Werte für die ARL des CU_t-Schemas bei h^{cu} = 3.5 unter Benutzung verschiedener Werte für N_{Gitter}. Die wahre Verteilung der X_t sei dabei eine Polya-Verteilung mit n = 125, $\bar{p}$ = 0.9 % und σ = 0.0 %, 0.2 %, ..., 2.2 %

	ARL bei versch. Werte von Ngitter					
σ in %	15	30	50	75	100	
0.0	38378	10324	19855	14807	14697	
0.2	16452	4959	8746	6699	6749	
0.4	2887	1104	1652	1348	1383	
0.6	604	295	385	338	345	
0.8	204	121	146	134	136	
1.0	101	69.6	79.2	75.4	76.0	
1.2	64.0	48.2	53.2	51.6	51.8	
1.4	46.9	37.6	40.7	39.9	39.9	
1.6	37.7	31.7	33.7	33.3	33.3	
1.8	32.4	28.1	29.6	29.4	29.3	
2.0	29.0	25.9	27.0	26.8	26.8	
2.2	27.0	24.4	25.3	25.2	25.1	

Tabelle 4.1 : Berechnung der ARL einer CUSUM-Karte aus Beispiel 4.1 fuer verschieden feine Unterteilungen des Intervalls $[0, h^{cu})$ und fuer verschiedene Werte der Prozessvarianz σ^2

Wir sehen, daß die Werte für die ARL in Abhängigkeit von N_{Gitter} oszillieren. Je größer der Wert der ARL, desto größer ist der Oszillationseffekt. Dieser unangenehme Effekt hat seine Ursache darin, daß die Verteilung von X_t und damit auch Y_t diskret ist und die Übergangsdichte stark davon abhängt, in welches Intervall Y_t fällt.

Für stetige Zuwachsverteilungen dürfen wir daher geringere Effekte der Intervalleinteilung auf die ARL erwarten. Brook und Ewans (1972) geben im Normalverteilungs-Fall $N_{Gitter} \approx 18$ als hinreichende Genauigkeit bei $h^{cu} = 3.0$ an.

In den Kapiteln 6 und 8 wird durchgängig $N_{Gitter} = 75$ zur Berechnung der ARL benutzt. □

Wir wollen zunächst die mehrstufigen Übergangswahrscheinlichkeiten zwischen den transienten Zuständen des verallgemeinerten CUSUM-Schemas berechnen.

Es sei hierzu μ die Summe von ν (vgl. (4.3)) und des auf -1 konzentrierten Zählmaßes. Der Träger von μ (Bez.: supp μ) ist dann:

$$\text{(4.16)} \qquad \text{supp } \mu := \begin{cases} -1,\ z_0 = 0,\ z_1,\ \ldots,\ z_{H-1} & \text{endlicher Fall} \\ -1 \cup [0,\ h^{cu}) & \text{kontinuierlicher Fall} \end{cases}$$

Bei gegebenen Übergangsdichten M_t $(t = 1, 2, \ldots)$ lassen sich die mehrstufigen Übergangsdichten M^n $(n = 1, 2, \ldots)$ auf den transienten Zuständen sukzessiv berechnen. Es gilt für $x, y, z \in \text{supp } \mu$

$$\text{(4.17)} \qquad M^1(x, y) = M_1(x, y)$$

$$\text{(4.18)} \qquad M^n(x, y) = \int M^{n-1}(x, z)\, M_n(z, y)\, \mu(dz) \qquad n \geq 2$$

die Wahrscheinlichkeit $P(N = n)$, daß das CUSUM-Schema genau n Schritte zum Erreichen des absorbierenden Zustands braucht, ist dann gegeben durch:

$$(4.19) \qquad P(N = 1) = 1 - \int M_1(o, y)\, \mu(dy)$$

$$(4.20) \qquad P(N = n) = \int M^{n-1}(o,x) \left(1 - \int M_n(x, y)\, \mu(dy)\right) \mu(dx) \qquad n \geq 2$$

Für großes n ist die Berechnung von $P(N = n)$ sehr aufwendig. Außerdem führt die rekursive Berechnung von M^n zu Einflüssen von Rundungsfehlern.

Abschnitt 4.2:

Das asymptotische Verhalten von M^n und $P(N = n)$ im stationären Fall

Für den stationären Fall, d. h. $M_t = M$ für alle $t \geq 1$, liefert das folgende Theorem eine Aussage über das asymptotische Verhalten von M^n:

Theorem 4.1:

M und μ mögen die folgenden Voraussetzungen erfüllen:

(*) Es existiert ein $n_o \in \mathbb{N}$ und $0<a<b<\infty$ mit:

$$a \leq M^{n_o}(x, y) \leq b \qquad x, y \in \text{supp } \mu$$

(**) $\int M(x, y)\mu(dy)$ ist eine beschränkte Funktion in x

(***) μ ist ein endliches Maß

Dann gelten die folgenden Existenzaussagen:

(i) M besitzt einen maximalen positiven Eigenwert $\lambda > 0$

(ii) Es existiert eine bis auf eine Konstante eindeutige rechte Eigenfunktion ϕ mit:

$$\phi(x) = \int M(x, y)\ \phi(y)\ \mu(dy) \qquad x \in \text{supp } \mu \tag{4.21}$$

und es existieren Schranken $0 < a_\phi < b_\phi < \infty$ mit

$$a_\phi \leq \phi(x) \leq b_\phi \qquad x \in \text{supp } \mu$$

(iii) Es existiert eine bis auf eine Konstante eindeutige linke Eigenfunktion ψ mit:

$$\psi(y) = \int \psi(x)\ M(x, y)\ \mu(dx) \qquad y \in \text{supp } \mu \tag{4.22}$$

und es existieren Schranken $0 < a_\psi < b_\psi < \infty$ mit:

$$a_\psi \leq \psi(y) \leq b_\psi \qquad y \in \text{supp } \mu$$

(iv) Es existiert ein von x und y unabhängiges $\Delta \in (0, 1)$, so daß unter der Normierung $\int \phi(x) \psi(x) \mu(dx) = 1$ gilt:

$$M^n(x, y) = \lambda^n \phi(x) \psi(y) (1 + 0(\Delta^n)) \qquad x, y \in \text{supp } \mu \tag{4.23}$$

Hierbei bezeichnet $0(\Delta^n)$ eine Zahlenfolge mit

$$\lim_{n \to \infty} \frac{0(\Delta^n)}{\Delta^n} = \text{const} \tag{4.24}$$

□

Der Beweis dieses Theorems ist bei Harris (1963), Appendix 3, zu finden.

Ist die Anzahl von transienten Zuständen endlich, kann M mit einer Matrix identifiziert werden. Theorem 4.1 ist in diesem Fall identisch mit dem klassischen Satz von Perron-Frobenius (vgl. Cox and Miller (1965), p. 118 ff).

Im vorliegenden Fall ist Bedingung (***) stets erfüllt. Wegen

$$\int M(x, y) \mu(dy) \leq 1$$

ist auch Bedingung (**) stets erfüllt.

Wir wollen die Bedingung (*) zunächst für den Fall untersuchen, daß die Verteilung der Zuwachsfunktionen $Y(s_i)$ $i = 1, \ldots, S$ diskret ist.

Bedingungen (*) ist sicher nicht erfüllt, falls ein $z \in \{z_{-1} = -1, z_o = 0, \ldots, z_{H-1}\}$ existiert mit $M^n(-1, z) = 0$ für alle $n \in \mathbb{N}$. Es bezeichne $\mathcal{Z}_{-1}$ die Menge aller Zustände, die von E_{-1} in endlichen vielen Schritten erreicht werden kann.

Satz 4.2:

Die Voraussetzung (*) in Theorem 4.1 ist erfüllt, falls:

$$\{E_{z_0}, E_{z_1}, \ldots, E_{z_{H-1}}\} \subset \mathbb{Z}_{-1} \tag{4.25}$$

$$P(Y(s_i) < 0) > 0 \text{ für alle } i = 1, \ldots, S \tag{4.26}$$

Beweis:

Wegen (4.26) gilt:

$$M^m(-1, -1) > 0 \text{ für jedes } m \in \mathbb{N}$$

Setze für $E_x \in \mathbb{Z}_{-1}$:

$$m_{-1,x} := \inf\{m \mid M^m(-1,x) > 0\}$$

$$m_{x,-1} := \inf\{m \mid M^m(x,-1) > 0\}$$

$m_{-1,x}$ ist nach Voraussetzung endlich. Wegen (4.26) ist auch $m_{x,-1}$ endlich. Setze:

$$h_o := \max_{E_x \in \mathbb{Z}_{-1}} m_{x,-1} + \max_{E_y \in \mathbb{Z}_{-1}} m_{-1,y}$$

Dann erfüllt M^{n_o} die Bedingung (*). Dies sieht man folgendermaßen ein:

Um von E_x in den Zustand E_{-1} zu kommen, werden $m_{x,o}$ Schritte benötigt. Danach bleibt das System $n_o - m_{x,-1} - m_{-1,y}$ Stufen mit positiver Wahrscheinlichkeit im Zustand E_{-1}. Der Zustand E_y wird dann mit $m_{-1,y}$ Schritten erreicht.

Die Übergänge von E_{-1} nach E_x und E_x nach E_{-1} behandelt man analog.

□

Bemerkung 4.2:

Bei einer einfachen CUSUM-Karte fällt $\bar{Z}_{-1}$ zusammen mit der Menge aller Zustände, die von E_o in endlich vielen Schritten erreicht werden können.

Ein Zustand, der vom Startzustand E_o nicht in endlich vielen Schritten erreicht werden kann, ist jedoch irrelevant und kann eliminiert werden.

□

Für den Fall, daß die Zuwachsfunktionen $Y(s_i)$ eine Dichte $g(y|s_i)$ $(i = 1, \ldots, S)$ bezüglich des Lebesgue-Maßes λ besitzen, erhält man:

Satz 4.3:

Die Voraussetzung (*) im Theorem 4.1 ist erfüllt, falls die $g(y|s_i)$ $(i = 1, \ldots, S)$ auf $[-h^{cu}, h^{cu}]$ beschränkt sind und ein $\varepsilon > 0$ und ein $y_\varepsilon > 0$ existieren mit:

(4.27)
$$\bigwedge_{y \in [-y_\varepsilon,\, y_\varepsilon]} \quad \bigwedge_{i = 1, \ldots, S} \quad g(y|s_i) \geq \varepsilon$$

Beweis:

Aus (4.7) und (4.27) folgt:

$$M^m(-1, -1) \geq (\varepsilon y_\varepsilon)^m \qquad m \in \mathbb{N}$$

Definiere $g(y)$ als das Minimum der $g(y|s_i)$ $i = 1, \ldots, S$. Es folgt aus (4.8) und (4.27) für $y \in [o, h^{cu})$:

$$M^1(-1, y) \geq g(y) \geq \varepsilon \qquad \text{für} \quad 0 < y \leq y_\varepsilon$$

$$M^2(-1, y) = \int M^1(-1, x)\, M(x, y)\, \mu(dx)$$

$$\geq \int_{\frac{1}{2} y_\varepsilon}^{y_\varepsilon} \varepsilon\, g(y - x)\, \mu(dx)$$

$$\geq \varepsilon^2 \frac{y_\varepsilon}{2} \qquad \text{für} \quad y \in \left[\frac{2 y_\varepsilon}{2}, \frac{3 y_\varepsilon}{2}\right]$$

Durch fortgesetzte Anwendung dieses Schritts erhalten wir:

$$M^m(-1,y) \geq \varepsilon \left(\frac{\varepsilon y_\varepsilon}{2}\right)^{m-1} \quad \text{für} \quad y \in \left[n\frac{y_\varepsilon}{2}, (n+1)\frac{y_\varepsilon}{2}\right]$$

Es sei $U_1 := [o, y_\varepsilon)$ und $U_n := [n\frac{y_\varepsilon}{2}, (n+1)\frac{y_\varepsilon}{2})$ für $n \geq 2$. Dann läßt sich jedem $y \in [o, h^{cu})$ genau ein $m_{-1,y}$ zuordnen, so daß $y \in U_{m_{-1,y}}$ gilt.

Analog können wir Abschätzungen für $M^n(x,-1)$ finden. Aufgrund der Symmetrie gilt, daß $M^n(x,-1) \geq \varepsilon (\frac{\varepsilon y_\varepsilon}{2})^{n-1}$ ist für $x \in U_n$. Wiederum läßt sich zu jedem $x \in [o, h^{cu})$ ein $m_{x,-1}$ zuordnen, so daß $x \in U_{m_{x,-1}}$. Aufgrund der Symmetrie gilt $m_{x,-1} = m_{-1,x}$.

Setze $$h_o := 2 \max_{x \in [0, h^{cu})} m_{-1,x}$$

Dann erfüllt M^{n_o} die Voraussetzung (*) von Theorem 4.1, denn für $x,y \in [o, h^{cu})$ gilt mit $n_{diff} := n_o - m_{x,-1} - m_{-1,y}$:

$$M^{n_o}(x,y) \geq M^{m_{x,-1}}(x,-1)\, M^{n_{diff}}(-1,-1)\, M^{m_{-1,y}}(-1,y)$$

$$\geq \varepsilon \left(\frac{\varepsilon y_\varepsilon}{2}\right)^{h_o-1} \cdot (\varepsilon y_\varepsilon)^{h_o} \cdot \varepsilon \left(\frac{\varepsilon y_\varepsilon}{2}\right)^{h_o-1}$$

Analog lassen sich $M^{n_o}(-1,y)$ und $M^{n_o}(x,-1)$ unabhängig von x und y abschätzen.

Die Existenz einer oberen Schranke für $M^{n_o}(x,y)$ folgt sofort aus der Beschränktheit der $g(y|s_i)$ $i = 1, \ldots, S$.

□

Falls die X_t ($t = 1, 2, \ldots$) aus einer 1-parametrigen Exponentialfamilie stammen, sind die Zuwachsfunktionen $Y(s_i)$ proportional zu $x - k(s_i)$ (vgl. (2.7)). Bedingung (4.27) ist erfüllt, falls $f(x|s_i, \vartheta)$ an der Stelle $x = k(s_i)$ ($i = 1, \ldots, S$) stetig und echt positiv ist. Die Bedingung (4.26) ist erfüllt, falls:

(4.28) $\quad P(X < k(s_i)) > 0 \quad$ für alle $i = 1, \ldots, S$

Mit Hilfe von Theorem 4.1 können wir auf das ergodische Verhalten von $(GCU_t)_{t \geq 1}$ auf den transienten Zuständen schließen. Wir definieren hierzu:

$$(4.29) \qquad m^n(x,z) := \frac{M^n(x,z)}{\int M^n(x,z)\,\mu(dz)}$$

$m^n(x,z)$ ist die bedingte Dichte von $GCU_{t=n}$ auf den transienten Zuständen, wenn das CUSUM-Verfahren bei $GCU_{t=0} = x$ gestartet wurde und bis zum Zeitpunkt t=n noch nicht die Alarmgrenze überschritten hat. Der folgende Satz besagt, daß $m^n(x,z)$ gegen einen vom Startwert x unabhängigen Grenzwert konvergiert.

Satz 4.4:

Unter den Voraussetzungen des Theorems 4.1 gilt für alle $x \in$ supp μ:

$$(4.30) \qquad m^n(x,z) \xrightarrow[n \to \infty]{} \frac{\psi(z)}{\int \psi(z)\,\mu(dz)}$$

Beweis:

Aus $M^n(x,z) = \lambda^n \phi(x) \psi(z) \; (1 + 0(\Delta^n))$ mit $0 < \Delta < 1$ und der Beschränktheit von $\phi(x)$ und $\psi(z)$ folgt die gleichmäßige Konvergenz von $M^n(x,z)$ gegen $\lambda^n \phi(x) \psi(z)$. Hieraus folgt:

$$m^n(x,z) = \frac{\lambda^n \phi(x)\psi(z)\,(1+0(\Delta^n))}{\int \lambda^n \phi(x)\psi(z)\,(1+0(\Delta^n))\,\mu(dz)}$$

$$\xrightarrow[n \to \infty]{} \frac{\psi(z)}{\int \psi(z)\,\mu(dz)} \qquad \square$$

Mit Hilfe von Theorem 4.1 läßt sich das asymptotische Verhalten von $P(N = n)$ einfach beschreiben.

Satz 4.5:

Unter den Voraussetzungen des Theorems 4.1 gilt:

$$(4.31) \quad P(N=n) \xrightarrow[n \to \infty]{} \lambda^{n-1}(1-\lambda)\,\phi(0) \int \psi(y)\,\mu(dy)$$

Beweis:

$$P(N=n) = \int M^{n-1}(0,x)\left(1 - \int M(x,y)\,\mu(dy)\right)\mu(dx)$$

$$\xrightarrow[n \to \infty]{} \lambda^{n-1}\phi(0) \int \psi(x)\left(1 - \int M(x,y)\,\mu(dy)\right)\mu(dx)$$

Aufgrund der Eigenwertgleichung für $\psi(x)$ erhält man:

$$\int \psi(x)\left(1 - \int M(x,y)\,\mu(dy)\right)\mu(dx)$$

$$= \int \psi(x)\,\mu(dx) - \iint \psi(x)\,M(x,y)\,\mu(dx)\,\mu(dy)$$

$$= (1-\lambda) \int \psi(x)\,\mu(dx) \qquad \square$$

Satz 4.5 verallgemeinert die Resultate von Brook and Evans (1972).

Die Verteilung der Lauflänge N gleicht für große n somit einer geometrischen Verteilung. Dabei ist $(1 - \lambda)$ die Wahrscheinlichkeit beim nächsten Schritt in den absorbierenden Zustand zu geraten.

Abschnitt 4.3:

Obere und untere Schranken für P(N), E(N), VAR(N) und N_α

In der Regel sind die Eigenfunktionen ϕ und ψ von M nur schwer zu ermitteln, so daß die explizite Darstellung des Grenzwertes von P(N = n) wenig Nutzen bietet. Woodall (1983) hat vorgeschlagen, die P(N = n) so lange rekursiv über (4.20) zu berechnen, bis sich

$$\frac{P(N = n-1)}{P(N = n)} \approx \lambda$$

stabilisiert hat und die übrigen Werte durch

(4.32) $\quad P(N = n+i) \approx \lambda^i P(N = n) \quad i \in \mathbb{N}$

zu approximieren. Ein solches Verfahren hat jedoch den Nachteil, keine Schranken für die Güte der Approximation zu liefern.

Wegen (4.32) ist es naheliegend, Schranken c_n^- , c_n^+ ($n \in \mathbb{N}$) zu suchen, für die gilt:

(4.33) $\quad (c_n^-)^i P(N = n) \leq P(N = n+i) \leq (c_n^+)^i P(N = n) \quad i \in \mathbb{N}$

Wegen

$$P(N = n) = \int M^{n-1}(o,x)\,(1 - \int M(x,y,)\mu(dy)\,)\mu(dx)$$

kann P(N = n) abgeschätzt werden, wenn $M^{n-1}(o,x)$ gleichmäßig gegen obere und untere Schranken abgeschätzt wird. Eine geeignete Wahl liefert die folgende Abschätzung:

$$M^{n+1}(o,y) = \int M^n(o,x)\, M(x,y)\, \mu(dx)$$

$$= \int \frac{M^n(o,x)}{M^{n-1}(o,x)}\, M^{n-1}(o,x)\, M(x,y)\, \mu(dx)$$

$$\leq \left(\sup_x \frac{M^n(o,x)}{M^{n-1}(o,x)} \right) M^m(o,y)$$

Die obige Abschätzung liefert den Ansatz:

(4.34) $$c_n^+ := \sup_x \frac{M^n(o,x)}{M^{n-1}(o,x)}$$

(4.35) $$c_n^- := \inf_x \frac{M^n(o,x)}{M^{n-1}(o,x)}$$

Hierbei läuft die Supremum- bzw. Infinimum-Bildung über alle $x \in \text{supp } \mu$.

Das folgende Lemma sichert die Existenz von c_n^+ ab einem gewissen n_o:

Lemma 4.6:
M erfülle die Bedingung (*) von Theorem 4.1 mit n_o, $0 < a_o < b_o < \infty$. Dann existieren für alle $n \geq n_o$ $0 < a_n < b_n < \infty$ mit:

(4.36) $$0 < a_n \leq M^n(x,y) \leq b_n < \infty \qquad x,y \in \text{supp } \mu$$

Beweis:

Wir zeigen die Behauptung für $n = n_o + 1$. Der Rest der Behauptung folgt dann über Induktionsschluß.

Nach Teil (iii) von Theorem 4.1 existiert eine linke Eigenfunktion ψ von M und Konstanten $0 < a_\psi < b_\psi < \infty$ mit :

$$a_\psi \leq \psi(y) \leq b_\psi \qquad y \in \text{supp } \mu$$

Es gilt:

$$M^{n_o+1}(x,y) = \int M^{n_o}(x,z) M(z,y) \mu(dz)$$

$$\leq \int M^{n_o}(x,z) \frac{\psi(z)}{a_\psi} M(z,y) \mu(dz)$$

$$\leq \frac{b_o}{a_\psi} \int \psi(z)\, M(z,y)\, \mu(dz)$$

$$= \frac{b_o}{a_\psi}\, \lambda\, \psi(y)$$

$$\leq \frac{b_o}{a_\psi}\, \lambda\, b_\psi$$

Die Existenz einer unteren Schranke für M^{n_o+1} folgt analog. □

Das folgende Theorem (vgl. Waldmann(1984 a) zeigt, daß die in (4.33) beschriebenen Schranken monoton gegen $P(N = n + i)$ konvergieren.

Theorem 4.7:

M erfülle die Bedingung (*) von Theorem 4.1 mit $n = n_o$. Dann gilt für alle $n > n_o$

(i) $(C_n^-)^i\, P(N = n) \leq P(N = n + i) \leq (C_n^+)^i\, P(N = n) \quad i \geq 1$

(ii) $(C_n^-)^i\, P(N = n) \leq (C_{n+1}^-)^{i-1}\, P(N = n + 1) \quad 1 \leq l \leq i$

(iii) $(C_n^+)^i\, P(N = n) \geq (C_{n+1}^+)^{i-1}\, P(N = n + 1) \quad 1 \leq l \leq i$

(iv) $\lim_{n\to\infty} c_n^+ = \lim_{n\to\infty} C_n^- = \lambda > 0$

□

Ein Beweis dieses Theorems ist bei Waldmann (1984 a) zu finden.

Die über c_n^- und c_n^+ gewonnenen Schranken für $P(N = n + i)$ $(i \geq 1)$ liefern Abschätzungen für E(N), VAR(N) und das α-Quantil der Lauflängenverteilung. Definiere zu diesem Zweck:

$$\hat{E}_\lambda(N) := \sum_{m=1}^{n} m\, P(N=m) + P(N=n)\, \lambda \left(\frac{n}{1-\lambda} + \frac{1}{(1-\lambda)^2} \right) \tag{3.37}$$

$$\hat{E}_\lambda(N^2) := \sum_{m=1}^{n} m^2 P(N=m) + P(N=n)\,\lambda\left(\frac{n^2}{1-\lambda} + \frac{2n}{(1-\lambda)^2} + \frac{1+\lambda}{(1-\lambda)^3}\right) \tag{3.38}$$

$$\hat{N}_\lambda(\alpha) := n + \frac{1}{\ln\lambda}\ln\left(1 + \frac{1-\lambda}{\lambda}\cdot\frac{\sum_{m=1}^{n} P(N=m) - \alpha}{P(N=n)}\right) \tag{3.39}$$

Definiere das α-Quantil der Lauflängenverteilung durch:

$$N_\alpha := \max\left\{ m \mid \sum_{k=1}^{m} P(N=k) \leq \alpha,\ m \in \mathbb{N}_0 \right\} \tag{4.40}$$

Da die Verteilung von N diskret ist, wird das Gleichheitszeichen in (4.40) im allgemeinen nicht immer angenommen.

Satz 4.8:

M erfülle die Bedingung (*) von Theorem 4.1. c_n^- und c_n^+ seien für $n > n_0$ durch (4.34) und (4.35) definiert. Dann gilt für $n > n_0$ und $c_n^+ < 1$:

(i) $\hat{E}_{c_n^-}(N) \leq E(N) \leq \hat{E}_{c_n^+}(N)$

(ii) $\hat{E}_{c_n^-}(N^2) - \hat{E}^2_{c_n^+}(N) \leq VAR(N) \leq \hat{E}_{c_n^+}(N^2) - \hat{E}^2_{c_n^-}(N)$

(iii) Für $\sum_{k=1}^{n} P(N=k) < \alpha$ gilt:

$$\left[\hat{N}_{c_n^+}(\alpha)\right] \leq N_\alpha \leq \left[\hat{N}_{c_n^-}(\alpha)\right] + 1$$

Hierbei bezeichnet $[r]$, $r \in \mathbb{R}$ den ganzzahligen Anteil von r.

Beweis:

Setze

$$p_m(\lambda) := \begin{cases} P(N=m) & m \leq n \\ \lambda^{m-n} P(N=n) & m > n \end{cases}$$

Dann gilt für $|\lambda| < 1$:

$$\sum_{m=1}^{\infty} m\, p_m(\lambda) = \sum_{m=1}^{n} m\, P(N=m) + P(N=n)\left((n+1)\lambda + (n+2)\lambda^2 + \dots\right)$$

$$= \hat{E}_\lambda(N)$$

Analog zeigt man:

$$\hat{E}_\lambda(N^2) = \sum_{m=1}^{\infty} m^2\, p_m(\lambda)$$

Die Aussagen unter (i) und (ii) folgen dann aus:

$$(4.41) \qquad p_m(c_n^-) \leq P(N=m) \leq p_m(c_n^+) \qquad m \geq 1$$

$\hat{N}_\lambda(\alpha)$ wurde so gewählt, daß die folgende Gleichung gilt:

$$\sum_{m=1}^{n} P(N=m) + P(N=n)\left(\lambda + \lambda^2 + \dots + \lambda^{\hat{N}_\lambda(\alpha)-n}\right) = \alpha$$

Damit gilt:

$$\sum_{m=1}^{[\hat{N}_{c_m^+}(\alpha)]} p_m(c_n^+) \leq \alpha < \sum_{m=1}^{[\hat{N}_{c_n^+}(\alpha)]+1} p_m(c_n^+)$$

Zusammen mit (4.41) folgt hieraus sofort:

$$[\,\hat{N}_{c_n^+}(\alpha)\,] \leq N_\alpha$$

Analog zeigt man den 2. Teil der Ungleichung in (iii).

□

Abschnitt 4.4:

Lösung der Operatorgleichung für die ARL-Funktion

Man kann E(N) jedoch direkt - ohne die gesamte Verteilung von N zu bestimmen - berechnen. Hierzu bezeichne für $x \in \text{supp } \mu$:

$$ARL(x) := E(N|GCU_0 = x)$$

ARL(x) bezeichnet also die erwartete Lauflänge, wenn das CUSUM-Schema im Punkt $x \in \text{supp } \mu$ gestartet wird. Lucas and Crosier (1982) bezeichnen einen solchen Anfang einer CUSUM-Prozedur als "Head-Start", der einem Start bei $x = 0$ vorzuziehen ist.

Die ARL-Funktion genügt der folgenden Operator-Gleichung:

Satz 4.9:

Es sei $x \in \text{supp } \mu$

$$(4.42) \qquad ARL(x) = 1 + \int M(x, y)\, ARL(y)\, \mu(dy)$$

Beweis:

Für $z \in \mathbb{R}$ gilt:

$$ARL(x|GCU_1 = z) = \begin{cases} 1 & z \geq h^{cu} \\ 1 + ARL(z) & z < h^{cu} \end{cases}$$

Damit erhält man:

$$ARL(x) = \int ARL(x|GCU_1 = z)\, dP(GCU_1 = z|GCU_0 = x)$$

$$= 1 + \int ARL(z)\, M(x,z)\, \mu(dz)$$

□

Im endlichen Fall läßt sich Gleichung (4.42) explizit lösen, vgl. Brook and Evans (1972). Man erhält:

$$(4.43) \qquad \begin{pmatrix} \vdots \\ \mathrm{ARL}(x) \\ \vdots \end{pmatrix} = (I - M)^{-1} \begin{pmatrix} \vdots \\ 1 \\ \vdots \end{pmatrix}$$

Der Operator T mit

$$T(v)\ (x) := 1 + \int M(x,y)\ \ v(y)\ \mu(dy)$$

ist isoton, da aus $v(x) \leq w(x)$ für alle $x \in \operatorname{supp} \mu$ stets $T(v)\ (x) \leq T(w)\ (x)$ auf supp μ folgt. Die Isotonie des Operators T gestattet die Konstruktion von effizienten Algorithmen zur Berechnung der ARL-Funktion.

Definiere hierzu für jede beschränkte Funktion v auf supp μ:

$$T^0(v) = v$$

$$T^n(v) = T\ (T^{n-1}(v)\ ) \qquad n \geq 1$$

Satz 4.10:

Unter der Bedingung (*) von Theorem 4.1 gilt für jede beschränkte Funktion v auf supp μ:

$$T^n(v)\ (x) \underset{n \to \infty}{\longrightarrow} \mathrm{ARL}(x) \qquad x \in \operatorname{supp} \mu$$

Beweis:

Wegen (iv) von Theorem 4.1 konvergiert $M^n(x,y)$ gleichmäßig gegen 0. Wegen $T(\mathrm{ARL})\ (x) = \mathrm{ARL}(x)$ erhält man:

$$|\mathrm{ARL}(x) - T^h(v)(x)| = |\ T^h(\mathrm{ARL})(x) - T^h(v)(x)\ |$$

$$= \left| \int M(x,y) \left(T^{h-1}(ARL)(y) - T^{h-1}(v)(y) \right) \mu(dy) \right|$$

$$\vdots$$

$$= \left| \int M^h(x,y) \left(ARL(y) - v(y) \right) \mu(dy) \right|$$

$$\leq \left(\sup_y |ARL(y) - v(y)| \right) \int M^h(x,y) \mu(dy)$$

$$\xrightarrow[h \to \infty]{} 0$$

□

Erfüllt eine auf supp μ beschränkte Funktion v(x) die Bedingung

(4.44) $\quad v(x) \leq T(v)(x) \qquad x \in \text{supp } \mu$

so folgt aus der Isotonie von T:

(4.45) $\quad v(x) \leq T(v)(x) \leq T^2(v)(x) \leq \ldots \leq \lim_{n \to \infty} T^n(v)(x) = ARL(x)$

Damit liefert jedes v(x) mit (4.44) eine Folge von unteren Schranken, die monoton gegen ARL(x) konvergiert. Analog garantiert

$$v(x) \geq T(v)(x) \qquad x \in \text{supp } \mu$$

eine Folge von oberen Schranken, die gegen ARL(x) konvergiert.

Ein verbessertes Konvergenzverhalten erhält man durch den folgenden Extrapolations-Ansatz von Waldmann (1983).

Setze $v_0 \equiv 0$ und $v_n = Tv_{n-1} \qquad n \geq 1$

Auf der Basis der Differenz zweier Iterationen

$d_n = v_n - v_{n-1}$ wird eine untere Schranke u_n der ARL-Funktion folgendermaßen definiert:

$$u_n(x) := v_{n-1}(x) + e_n^- \, d_{n-1}(x) \qquad x \in \operatorname{supp} \mu$$

Der Faktor e_n^- wird dabei so gewählt, daß die Bedingung (4.44) für u_n gerade noch erfüllt ist. Dies liefert:

$$\begin{aligned} 0 &\leq T(u_n)(x) - u_n(x) \\ &= T(v_{n-1} + e_n^- d_{n-1})(x) + (v_{n-1} + e_n^- d_{n-1})(x) \\ &= v_n(x) + e_n^- d_n(x) - v_{n-1}(x) - e_n^- d_{n-1}(x) \\ &= d_n(x) - e_n^- (d_{n-1}(x) - d_n(x)) \end{aligned}$$

Damit muß gelten:

$$e_n^- = \inf_{x \in D_n} \frac{d_n(x)}{d_{n-1}(x) - d_n(x)} \tag{4.46}$$

wobei $D_n := \{ x \mid d_{n-1}(x) - d_n(x) > 0,\ x \in \operatorname{supp} \mu \}$

und $\inf \emptyset := 0$.

Analog erhält man für eine obere Schranke o_n der Form:

$$o_n(x) := v_{n-1}(x) + e_n^+ \, d_{n-1}(x)$$

die Bedingung:

$$e_n^+ = \sup_x \frac{d_n(x)}{d_{n-1}(x) - d_n(x)} \tag{4.47}$$

falls $d_{n-1}(x) > d_n(x)$ für alle $x \in \operatorname{supp} \mu$.

Theorem 4.11:

Es sei $v_o \equiv 0$ und $v_n = Tv_{n-1}$ $(n \geq 1)$

Setze für $x \in \text{supp } \mu$:

$$U_n(x) := v_n(x) + e_n^- \, d_n(x)$$

$$O_n(x) := v_n(x) + e_n^+ \, d_n(x)$$

wobei e_n^- und e_n^+ durch (4.46) und (4.47) definiert sind.

Dann gilt für $x \in \text{supp } \mu$:

(i) $U_n(x) \leq U_{n+1}(x) \leq \ldots \leq ARL(x)$

(ii) $\lim\limits_{n \to \infty} U_n(x) = ARL(x)$

(iii) Falls $d_{n-1}(x) > d_n(x)$ für alle $x \in \text{supp } \mu$

$$O_n(x) \geq ARL(x)$$

(iv) Falls $d_{n-1}(x) > d_n(x) > d_{n+1}$ für alle $x \in \text{supp } \mu$

$$O_n(x) \geq O_{n+1}(x)$$

(v) Falls $d_{n-1}(x) > d_n(x)$ für alle $x \in \text{supp } \mu$ und für alle n ab einem gewissen n_o mit $e_{n_o}^+ < \infty$:

$$\lim_{n \to \infty} O_n(x) = ARL(x)$$

□

Ein Beweis dieses Theorems ist bei Waldmann (1983) zu finden.

Abschnitt 4.5:

Berechnung der Kenngrößen bei instationärer Verteilungsannahme

Bisher haben wir nur den Fall behandelt, daß die Übergangsdichte M_t zeitlich unabhängig ist, und damit die betrachtete Markov-Kette homogen ist.

Wir machen jetzt die etwas realistischere Annahme, daß der Parameter Θ_t zu einem unbekannten Zeitpunkt $t = \tau$ vom Sollzustand Θ_S in den Alarmzustand Θ_A springt. In diesem Fall ist

$$M_t = \begin{cases} M_S & t \leq \tau \\ M_A & t > \tau \end{cases}$$

Gemäß (4.20) erhalten wir:

$$(4.48)\quad P(N=n) = \int M_S^{n-1}(0,x)\left(1 - \int M_S(x,y)\,\mu(dy)\right)\mu(dx) \qquad 2 \leq n \leq \tau$$

$$(4.49)\quad P(N=n) = \int \pi(z) \int M_A^{n-\tau-1}(z,x)\left(1 - \int M_A(x,y)\,\mu(dy)\right)\mu(dx)\,\mu(dz) \qquad n > \tau$$

wobei $\pi(z) = M_S^{\tau}(0,z) \qquad (z \in \operatorname{supp} \mu)$

Die Verteilung von K_τ mit

$$K_\tau := \begin{cases} 0 & N \leq \tau \\ k & N = \tau + k \end{cases} \qquad k \in \mathbb{N}$$

ist dann gegeben durch:

$$P(K_\tau = k) = \begin{cases} P(N \leq \tau) & k = 0 \\ P(N = \tau + k) & k \geq 1 \end{cases}$$

Folglich:

$$P(K_\tau = 1) = \int \pi(z) \left(1 - \int M_A(z,y)\,\mu(dy)\right) \mu(dz)$$

$$P(K_\tau = k) = \int \pi(z) \int M_A^{k-1}(z,x) \left(1 - \int M_A(x,y)\,\mu(dy)\right) \mu(dx)\,\mu(dz) \qquad k \geq 2$$

Wir sehen, daß die Verteilung von K_τ derjenigen des stationären Modells unter M_A entspricht mit dem Unterschied, daß die Startverteilung auf den transienten Zuständen hier nicht auf 0 konzentriert ist, sondern durch $\pi(z)$, $z \in \text{supp}\ \mu$ gegeben ist.

Eine Reihe von Aussagen aus den Abschnitten 4.2 - 4.4 lassen sich daher auf die Verteilung von K_τ verallgemeinern. Es sei für $x, z \in \text{supp}\ \mu$

$$m_\tau^k(x,z) := \frac{M_A^k(x,z)}{\int M_A^k(x,z)\,\mu(dz)}$$

die bedingte Dichte von $GCU_{k+\tau}$ nach k Schritten auf den transienten Zuständen, wenn GCU_τ den Wert x hat und nach k weiteren Schritten die Alarmgrenze noch nicht überschritten hat.

Satz 4.12:

Falls M_A die Voraussetzung (*) von Theorem 4.1 erfüllt, gilt für alle $x \in \text{supp}\ \mu$:

(i) $$m_\tau^k(x,z) \xrightarrow[k \to \infty]{} \frac{\psi_A(z)}{\int \psi_A(z)\,\mu(dz)}$$

(ii) $$P(K_\tau = k) \xrightarrow[k \to \infty]{} \lambda_A^k (1 - \lambda_A) \int \phi_A(x)\,\pi(x)\,\mu(dx) \int \psi_A(y)\,\mu(dy)$$

Beweis:

(i) erhält man analog zu Satz 2.4.

Definiere $h_k(z)$ durch:

$$h_k(z) := \int M_A^{k-1}(z,x)\left(1 - \int M_A(x,y)\,\mu(dy)\right)\mu(dz)$$

Analog zu Satz 4.5 erhält man:

$$h_k(z) \xrightarrow[k\to\infty]{} \lambda_A^{k-1}\,(1-\lambda_A)\,\phi_A(z)\int \psi_A(y)\,\mu(dy)$$

Wegen $h_k(z) \leq 1$ für $z \in \operatorname{supp}\mu$ und $\int \pi(z)\,\mu(dz) \leq 1$ folgt (ii) nach dem Satz von der majorisierten Konvergenz. □

Für den Fall, daß wir obere und untere Schranken für $P(K_\tau = k)$ zu gegebenen τ berechnen wollen, können wir den zu (4.33) analogen Ansatz

$$(4.50)\qquad (\hat{c}_k^-)^i\,P(K=k) \leq P(K=k+i) \leq (\hat{c}_k^+)^i\,P(K=k) \quad i \geq 1$$

wählen, wobei:

$$(4.51)\qquad \hat{c}_k^- := \inf_{x,z} \frac{M_A^k(x,z)}{M_A^{k-1}(x,z)}$$

$$(4.52)\qquad \hat{c}_k^+ := \sup_{x,z} \frac{M_A^k(x,z)}{M_A^{k-1}(x,z)}$$

Gemäß Lemma 4.6 sind die $\hat{c}_k^{\pm}$ ab einem gewissen k_o wohl definiert und Theorem 4.7 kann entsprechend übernommen werden. Dies gilt auch für die in Satz 4.8 angegebenen Schranken.

Die Berechnung von $E(K_\tau)$ erfolgt gemäß:

$$E(K_\tau) = \int \pi(z)\, E(K_\tau | GCU_\tau = z)\, \mu(dz) \tag{4.53}$$

Für $ARL_{K_\tau}(z) := E(K_\tau | GCU_\tau = z)$ gilt die Operatorgleichung unter (4.42). Theorem 4.11 besagt, daß die Schranken der Form

$$v_k(x) + \hat{e}_k^- d_k(x) \leq ARL_{K_\tau}(x) \leq v_k(x) + \hat{e}_k^+ d_k(x) \tag{4.54}$$

monoton gegen $ARL_{K_\tau}(x)$ konvergieren, wobei $\hat{e}_k^-$ bzw. $\hat{e}_k^+$ analog zu (4.46) bzw. (4.47) definiert sind. Die Monotonie der Schranken in (4.54) sichert:

$$\lim_{k\to\infty} \left(\int v_k(x)\pi(x)\mu(dx) + \hat{e}_k^- \int d_k(x)\pi(x)\mu(dx) \right) \tag{4.55}$$
$$= E(K_\tau) =$$
$$\lim_{k\to\infty} \left(\int v_k(x)\pi(x)\mu(dx) + \hat{e}_k^+ \int d_k(x)\pi(x)\mu(dx) \right)$$

Der Erwartungswert der Gesamtlauflänge N berechnet sich bei gegebenen τ dann folgendermaßen:

$$E(N) = \sum_{n=1}^{\tau} n\, P(N=n) + \sum_{k=1}^{\infty} (\tau+k)\, P(K_\tau = k) \tag{4.56}$$
$$= \sum_{n=1}^{\tau} n\, P(N=n) + \tau\, P[N>\tau] + E(K_\tau)$$

Im endlichen Fall läßt sich E(N) explizit angeben (vgl. Abel und Fuchs (1984)):

Satz 4.13:

Es sei I die HxH Einheitsmatrix und 1 der Vektor, der aus H Einsen besteht.
Es gilt:

$$(4.57)\quad \begin{bmatrix} \\ ARL_N(x) \\ \\ \end{bmatrix} = (I-M_S)^{-1}(I-M_S)^{\tau}\,1 + M_S^{\tau}(I-M_A)^{-1}1 \qquad \square$$

Einen Beweis von (4.57) findet man bei Abel und Fuchs (1984)).

Ein Gütekriterium für CUSUM-Karten ist es, daß $E(K_\tau|N>\tau)$ für alle $\tau>0$ klein sein soll (vgl. Kapitel 3). Der folgende Satz sichert die Konvergenz von $E(K_\tau|N>\tau)$ für $\tau\to\infty$.

Satz 4.14:

Falls M_S die voraussetzung (*) von Theorem 4.1 erfüllt, gilt:

$$E(K_\tau|N>\tau) \xrightarrow[\tau\to\infty]{} \frac{\int ARL_{K_\tau}(z)\,\psi_S(z)\,\mu(dz)}{\int \psi_S(z)\,\mu(dz)}$$

Beweis:

Es gilt:

$$E(K_\tau|N>\tau) = \frac{\int ARL_{K_\tau}(z)\,M_S^{\tau}(0,z)\,\mu(dz)}{\int M_S^{\tau}(0,z)\,\mu(dz)}$$

Damit folgt aus (4.23):

$$(4.58)\quad E(K_\tau|N>\tau) = \frac{\int ARL_{K_\tau}(z)\,\psi_S(z)\,(1+O(\Delta^\tau))\,\mu(dz)}{\int \psi_S(z)\,(1+O(\Delta^\tau))\,\mu(dz)}$$

Da $ARL_{K_\tau}(z)$ und $\psi_S(z)$ auf supp μ beschränkt sind, folgt die Konvergenz der Integrale im Zähler und im Nenner von (4.58).

$\square$

Kapitel 5

Das simultane Führen mehrerer CUSUM-Karten

Abschnitt 5.1:

Darstellung des Verfahrens als Markov-Kette mit mehrdimensionalem Zustandsraum

Im Kapitel 7 wird eine Situation behandelt, in der mehrere (einfache) CUSUM-Karten gleichzeitig geführt werden. Ein Alarm wird dann gegeben, wenn eine der Karten die Alarmgrenze überschreitet.

Formal läßt sich dieses simultane Führen mehrerer CUSUM-Karten folgendermaßen beschreiben: Es sei

$$Y_t(i) := \ln \frac{f(X_t|\,\sigma_{A_i})}{f(X_t|\,\sigma_S)} \qquad i = 1,2,\ldots,N_A \tag{5.1}$$

der Zuwachs der i-ten CUSUM-Karte zum Zeitpunkt t. Dabei ist $f(x|\,\sigma_{A_i})$ die Dichte der beobachteten Zufallsgrößen X_t unter der i-ten Alternative σ_{A_i} (i = 1, 2, ..., N_A). Die Alarmgrenzen der einzelnen CUSUM-Karten seien mit h(i) (i = 1, 2, ..., N_A) bezeichnet. Dann gilt:

$$CU_o(1) = \ldots = CU_o(N_A) = 0$$

(5.2)

$$CU_t(i) = \max \left\{ 0,\ CU_{t-1}(i) + Y_t(i) \right\} \qquad t \geq 1$$
$$i = 1, 2, \ldots, N_A$$

Alarm wird gegeben, falls:

(5.3) $\quad CU_t(i) \geq h(i)$ für ein $i \in \{1, 2, \ldots, N_A\} \quad t \geq 1$

Es bezeichne N(i) die Lauflänge der i-ten CUSUM-Karte, wobei:

$$N(i) = \inf \left\{ t \mid CU_t(i) \geq h(i),\ t \in \mathbb{N} \right\}$$
$$i = 1, 2, \ldots, N_A$$

Die Lauflänge des Gesamtsystems ist bestimmt durch:

$$N := \min \left\{ N(1), \ldots, N(N_A) \right\}$$

Unter der Bezeichnung

$$\bar{Y}_t := (Y_t(1), \ldots, Y_t(N_A))^T$$

$$\bar{\bar{CU}}_t := (CU_t(1), \ldots, CU_t(N_A))^T$$

$$\bar{h} := (h(1), \ldots, h(N_A))^T$$

$$\bar{o} := (o, \ldots, o)^T$$

erhält man die zu (5.2) und (5.3) äquivalente Beschreibung:

$$\bar{\bar{CU}}_o = \bar{0}_o$$

(5.4) $\quad \bar{\bar{CU}}_t = \max \left\{ \bar{o},\ \bar{\bar{CU}}_{t-1} + \bar{Y}_t \right\} \quad t \geq 1$

(5.5) Kein Alarm, solange $\bar{\bar{CU}}_t < \bar{h}$

Hierbei ist die Funktion max in (5.4) komponentenweise anzuwenden und die Relation "$<$" in (5.5) ist erfüllt, wenn sie für alle Komponenten gilt.

Wie im eindimensionalen Fall bilden die $(\overline{CU}_t)_{t\geq 0}$ eine Markov-Kette mit einem absorbierenden Zustand. Die transienten Zustände des Gesamtsystems lassen sich als kartesisches Produkt der transienten Zustände der einzelnen CUSUM-Karten darstellen.

Im eindimensionalen Fall konnte man aufgrund der Rekursionsbeziehung (5.4) sofort eine Übergangsdichte M der zugehörigen Markov-Kette $(\overline{CU}_t)_{t\geq 0}$ angeben. Im mehrdimensionalen Fall ist dies nur noch bei diskretem Zuwachsvektor $\overline{Y}$ ohne weiteres möglich.

Es sei μ das auf

$$\overline{Z} := \bigtimes_{i=1}^{N_A} \{z_0(i), z_1(i), \ldots, z_{H(i)-1}(i)\}$$

konzentrierte Zählmaß, wobei

$\{z_0(i) = 0, z_1(i), \ldots, z_{H(i)-1}(i)\}$ die Menge der Werte ist, die von der i-ten CUSUM-Karte angenommen werden können.
Dann gilt:

$$M_t(\overline{CU}_{t-1} = \overline{x}, \overline{CU}_t = \overline{v}) = P\Big(\bigcap_{i:\, v(i)=0} \{Y_t(i) \leq -x(i)\} \cap \bigcap_{j:\, v(j)>0} \{Y_t(j) = v(j) - x(j)\}\Big) \tag{5.6}$$

M_t genügt damit der Gleichung:

$$\bigwedge_{\overline{x} \in \overline{Z}} \bigwedge_{\overline{A} \subset \overline{Z}} P(\overline{CU}_t \in \overline{A} \mid \overline{CU}_{t-1} = \overline{x}) = \int_{\overline{A}} M_t(\overline{x}, \overline{v})\, \mu(d\overline{v}) \tag{5.7}$$

Bemerkung 5.1:

Im kontinuierlichen Fall versagt diese einfache Übertragung vom eindimensionalen Fall auf den mehrdimensionalen Fall. Dies hat seinen Grund darin, daß der Zuwachsvektor $\overline{Y}_t$ bezüglich des mehrdimensionalen Lebesgue-Maßes λ^{N_A} auf einer Nullmenge konzentriert ist.

Ein adäquater Ansatz zur Berechnung der ARL-Funktion für diesen Fall wird im Abschnitt 5.3 gegeben.

□

Abschnitt 5.2:

Berechnung der Lauflängenverteilung und der ARL des Gesamtsystems im diskreten Fall

Für den diskreten Fall wollen wir jetzt untersuchen, welche der Ergebnisse aus Kapitel 4 auf den mehrdimensionalen Fall übertragbar sind.

Für kleines n lassen sich die n-stufigen Übergangsdichten $M^n(\bar{x}, \bar{z})$ wieder rekursiv berechnen. Man erhält dann für $P(N = n)$:

$$P(N = 1) = 1 - \int M_1(\bar{o}, \bar{z})\ \mu(d\bar{z}) \tag{5.8}$$

$$P(N = n) = \int M^{n-1}(\bar{o}, \bar{x})\left(1 - \int M_n(\bar{x}, \bar{y})\ \mu(d\bar{y})\right)\mu(d\bar{x}) \qquad n \geq 2 \tag{5.9}$$

Um asymptotische Aussagen über das Theorem 4.1 machen zu können, müssen wir die Voraussetzung (*) von Theorem 4.1 überprüfen.

Im allgemeinen werden nicht alle Werte aus $\bar{Z}$ von den $\overline{CU}_t$ $(t \geq 0)$ angenommen. Es sei daher $\bar{Z}_{\bar{o}}$ die Menge aller Werte, die von den $\overline{CU}_t$ $(t \geq 0)$ nach endlich vielen Schritten von Startzustand $\bar{o}$ erreicht werden können.

Die Reduktion von $\bar{Z}$ auf $\bar{Z}_{\bar{o}}$ kann zwar die Menge der betrachteten transienten Zustände stark reduzieren, allerdings kann die Angabe von $\bar{Z}_{\bar{o}}$ mit einem erheblichen Rechenaufwand verbunden sein.

Analog zum Beweis von Satz 4.2 läßt sich die Voraussetzung (*) von Theorem 4.1 zeigen, falls die beiden folgenden Bedingungen erfüllt sind.

Bedingung 5.1:

Für jedes $\bar{z} \in \bar{Z}_{\bar{o}}$ ist die Wahrscheinlichkeit für einen Übergang von $\bar{z}$ nach $\bar{o}$ nach endlich vielen Schritten positiv. □

Bedingung 5.2:

Die Wahrscheinlichkeit für ein Verbleiben im Zustand $\bar{o}$ ist positiv. □

Für den allgemeinen Fall lassen sich die Bedingungen 5.1 und 5.2 nicht verifizieren. Bedingung 5.2 ist z. B. verletzt, falls:

$$P\Big(\bigcap_{i=1}^{N_A} \{Y(i) \leq 0\}\Big) = 0 \tag{5.10}$$

Beispiel 5.1:

Die X_t seien poisson-verteilt mit dem Parameter λ. Für $\lambda_{A_1} > \lambda_S > \lambda_{A_2}$ erhält man gemäß Satz 2.2 die zu den CUSUM-Karten äquivalenten Karten K_t^+ und K_t^-:

$$K_o^+ = K_o^- = 0$$

$$K_t^+ = \max\{0,\ K_{t-1}^+ + X_t - k^+\} \qquad t \geq 1 \tag{5.11}$$

$$K_t^- = \min\{0,\ K_{t-1}^- + X_t + k^-\} \qquad t \geq 1 \tag{5.12}$$

wobei k^+ und k^- gegeben sind durch :

(5.13) $$k^+ = \frac{\lambda_{A_1} - \lambda_S}{\ln \lambda_{A_1} - \ln \lambda_S}$$

(5.14) $$k^- = \frac{\lambda_{A_2} - \lambda_S}{\ln \lambda_S - \ln \lambda_{A_2}}$$

Es gelte $\lambda_{A_1} = f^+ \lambda_S$ ($f^+ > 1$) und $\lambda_{A_2} = f^- \lambda_S$ ($f^- < 1$). Dies liefert:

$$k^+ = \frac{f^+ - 1}{\ln f^+} \lambda_S \qquad k^- = \frac{1 - f^-}{\ln f^-} \lambda_S$$

Für $\lambda_S = 2$ und $k^+ = 3$ und $k^- = -2$ erhält man $f^+ \approx 2.2$ und $f^- \approx 0.45$. Folglich $\lambda_{A_1} \approx 4.4$ und $\lambda_{A_2} \approx 0.9$.

Für $h^{K+} = h^{K-} = 5.0$ sind die Punkte, die von (K_t^+, K_t^-) $(t \geq 0)$ angenommen werden können, in Abbildung 5.1 dargestellt.

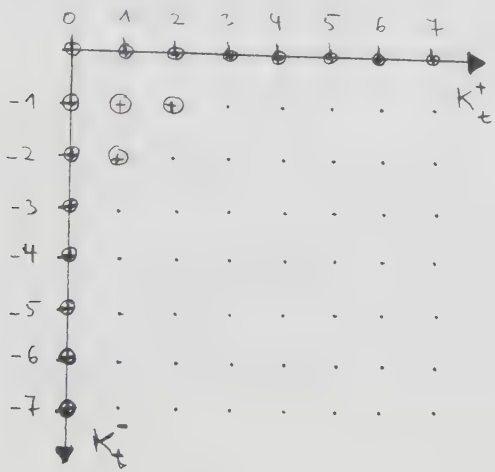

Abb. 5.1 : Darstellung aller Punkte (i,j) , die von (K_t^+, K_t^-) erreicht werden koennen

Wir sehen, daß die Anzahl der transienten Zustände von 25 auf 12 verringert werden kann. Die Übergangsmatrix M reduziert sich also von 625 auf 124 Elemente!

Die Übergangsmatrix zwischen diesen 12 Zuständen erfüllt die Bedingungen 5.1 und 5.2, denn:

Von jedem der 12 Zustände kann der Punkt (4,0) erreicht werden. Der Übergang von (4,0) zu (0,0) kann z. B. erfolgen über:

$$(4,0) \xrightarrow{x=0} (1,-2) \xrightarrow{x=3} (1,-1) \xrightarrow{x=3} (1,0) \xrightarrow{x=2} (0,0)$$

Die Wahrscheinlichkeit, daß (K_t^+ , K_t^-) in (0,0) verweilt, ist durch $P(X = 2) + P(X = 3) > 0$ gegeben. □

Bemerkung 5.2:

Woodall (1984) hat den Fall, daß gleichzeitig 2 CUSUM-Karten der Form (5.11) und (5.12) geführt werden, allgemein behandelt. Unter den Bedingungen

(5.13) $\quad P(X_t - k^+ \neq X_t + k^-) > 0 \qquad t = 1, 2, \ldots$

(5.14) $\quad P(X_t - k^+ = z) > 0$ und $P(X_t + k^- = z) > 0$

für alle z mit $|z| \leq \max \{ h^{K_+}, h^{K_-} \}$

hat Woodall den Wertebereich von (K_t^+ , K_t^-) bestimmt und bewiesen, daß die Bedingungen 5.1 und 5.2 erfüllt sind.

$\bar{Z}_0$ ist durch die folgenden Wertepaare gegeben:

(5.15)
$$(i_+, 0) \qquad i_+ = 0, \ldots, h^{K+} - 1$$
$$(0, i_-) \qquad i_- = -1, \ldots, -h^{K-} + 1$$

und

$$(i_+, i_-) \quad \text{mit} \quad 0 \leq i_+ \leq h^{K+}$$
$$0 \geq i_- \geq -h^{K-}$$

$$2 \leq i_+ - i_- \leq \max \{ h^{K+}, h^{K-} \} - (k^+ + k^-)$$

Der in Abbildung 5.1 angegebene Wertebereich stimmt mit dem in (5.15) und (5.16) definierten Bereich überein. Jedoch wird in Beispiel 5.1 die Voraussetzung (5.14) nicht erfüllt, denn es gilt mit den Werten aus Beispiel 5.1: $P(X_t - 3 = -5) = 0$. □

Ist der Nachweis für das Erfülltsein der Bedingungen 5.1 und 5.2 erbracht, so können die Ergebnisse von Kapitel 4 ohne weiteres auf den mehrdimensionalen Fall übertragen werden. Insbesondere können Schranken für $P(N = n)$ (Satz 4.7), $E(N)$, $VAR(N)$ und N_α (Satz 4.8) angegeben werden.

Die Lösung der Operatorgleichung

(5.17)
$$ARL(\bar{z}) = 1 + \int M(\bar{z}, \bar{y})\, ARL(\bar{y})\, \mu(d\bar{y})$$

läßt sich wieder über den Extrapolationsansatz von Waldmann (1983) bestimmen (Theorem 4.11).

Ebenso können die Ergebnisse bezüglich der instationären Verteilungsannahme verallgemeinert werden (Satz 4.12, 4.13, 4.14).

Abschnitt 5.3:

Berechnung der Gesamt-ARL im kontinuierlichen Fall

Wie wir im Abschnitt 5.1 erwähnt haben, läßt sich bei absolut stetigen Zufallsgrößen X_t $(t \geq 1)$ der Ansatz über die Übergangsdichte der Markov-Kette $(CU_t)_{t \geq 0}$ nicht ohne weiteres auf den mehrdimensionalen Fall übertragen.

Die Operatorgleichung (5.17) für die ARL-Funktion läßt sich jedoch auch ohne die Benutzung einer Übergangsdichte M formulieren.

Definiere hierzu die Abbildung CU durch:

$$CU(\bar{z}, \bar{y}) := \begin{pmatrix} \vdots \\ \max\{0,\ z(i) + y(i)\} \\ \vdots \end{pmatrix}$$

wobei $\bar{z}$, $\bar{y}$ und $CU(\bar{z}, \bar{y})$ Elemente von $\mathbb{R}^{N_A}$ sind.

Weiterhin sei $1_{\bar{Z}}$ die Indikatorfunktion auf $\bar{Z} = [0, h(1)) \times \ldots \times [0, h(N_A))$.

Man erhält dann für $ARL(\bar{z})$ $(\bar{z} \in \bar{Z})$ die Operatorgleichung:

$$(5.18)\quad ARL(\bar{z}) = 1 + \int 1_{\bar{Z}}(CU(\bar{z}, \bar{y}(x)))\, ARL(CU(\bar{z}, \bar{y}(x)))\, f(x|\vartheta)\, d(x)$$

Im Vergleich mit der Operatorgleichung (5.17) wird hier das Übergangsgesetz zwischen den transienten Zuständen direkt mitformuliert.

Der Operator W mit:

$$W(v)(\bar{z}) := 1 + \int 1_{\bar{Z}}(CU(\bar{z}, \bar{y}(x)))\, v(CU(\bar{z}, \bar{y}(x)))\, f(x|\vartheta)\, d(x)$$

ist isoton. Es läßt sich zeigen (vgl. Waldmann (1984 b)), daß der Operator W einen Fixpunkt v* besitzt mit der folgenden Eigenschaft:

(5.19) $\quad ARL(\bar{o}) = v^*(\bar{o})$

und

(5.20) $\quad v^* = \lim_{n \to \infty} v_n$

wobei $v_n := W(v_{n-1})$ und $v_o \equiv 0$.

Der Extrapolations-Algorithmus von Theorem 4.11 ist dann wieder anwendbar und liefert obere und untere Schranken für die ARL des Gesamtsystems.

Bemerkung 5.3:

Für den Fall, daß die X_t $(t \geq 1)$ aus einer 1-parametrigen Exponentialfamilie stammen und die zugehörigen K_t^{+-} und K_t^{-}-Kontrollkarten symmetrisch sind, d. h. es gilt $k^+ = k^- = k$ und $h^{K+} = h^{K-} = h$, läßt sich die ARL des Gesamtsystems aus den erwarteten Lauflängen $E(N_+)$ und $E(N_-)$ der einzelnen CUSUM-Karten berechnen. Man erhält (vgl. Khan (1981)):

(5.21) $$E(N) = \frac{E(N_+)\,E(N_-)}{E(N_+) + E(N_-)}$$

□

Kapitel 6

CUSUM-Karten zur Kontrolle der Prozeßkurve

Abschnitt 6.1:

Modelle für die Stichprobenverteilung

Bei der Wareneingangs- bzw. Warenausgangsprüfung werden den ein- bzw. herausgehenden Warenlosen, die aus je N_L Einzelstücken bestehen, Stichproben vom Umfang n entnommen. Die entnommenen Einzelstücke werden darauf geprüft, ob sie eine verlangte Qualitätsnorm erfüllen oder nicht (attributive Kontrolle). Ist die Anzahl $X^S = X^S(N_L, n)$ der schlechten Stücke in der Stichprobe größer als eine vorgegebene Annahmezahl c, so wird das Warenlos beim Wareneingang zurückgewiesen oder beim Warenausgang total inspiziert.

Um die statistischen Eigenschaften eines solchen Verfahrens zu ermitteln, müssen Verteilungsannahmen über $X^S = X^S(N_L, n)$ gemacht werden. Hald (1981), S. 126, hat hierzu das folgende Modell aufgestellt:

(M I) Produktionsphase

Die Wahrscheinlichkeit, daß ein schlechtes Einzelstück hergestellt wird, betrage p. Der Anteil der schlechten Stücke an der gesamten Produktion betrage daher p.

(M II) Losformierung:

Die Formierung der Warenlose aus der Gesamtproduktion erfolge zufällig, so daß die Anzahl der schlechten Stücke $X^L = X^L(N_L)$ in einem Los bei gegebenen p binomialverteilt ist:

$$X^L(N_L)|p \sim B(N_L, p)$$

(M III) Stichprobenziehung:

Dem Warenlos wird eine Stichprobe entnommen (Ziehen ohne Zurücklegen). Bei gegebener Anzahl schlechter Stücke X^L in Los ist $X^S = X^{S(N_L, n)}$ hypergeometrisch verteilt:

$$X^S(n, N_L)|X^L \sim H(n, N_L, X^L)$$

□

Häufig unterliegt der Produktionsprozeß gewissen Schwankungen und die Wahrscheinlichkeit, mit der schlechte Stücke produziert werden, schwankt gemäß einer Verteilung W. Wenn W eine Dichte w(p) auf (0,1) hat, so heißt w die Prozeßkurve des Produktionsprozesses.

Ist W z. B. eine 2-Punkt-Verteilung mit Masse w_1 auf p_1 und Masse w_2 auf p_2, so modelliert dies eine Situation, in der mit 2 verschiedenen Fertigungstechniken produziert wird. Dabei ist die 1. Fertigungstechnik mit $100 \cdot w_1$ % am Gesamtproduktionsaufkommen beteiligt.

Die Verteilung der schlechten Stücke $X^L(N_L)$ in einem Los ist dann durch eine bzgl. W gemischte Binomialverteilung gegeben:

$$P\,(X^L(N_L) = X\,) = \int_0^1 b(X|N_L, p)\, dW(p)$$

Die Verteilung der schlechten Stücke $X^S = X^S(n, N_L)$ in der Stichprobe ist dann durch eine gemischte hypergeometrische Verteilung gegeben:

$$P(X^S(n,N_L)=x) = \sum_{X=0}^{N_L} h(x|n,X,N_L)\, P(X^L(N_L)=X)$$

Die Eigenschaften der gemischten hypergeometrischen Verteilung wurden von Hald (1960) ausführlich untersucht.

Interessant ist die Tatsache, daß die Stichprobenverteilung die Verteilung der schlechten Stücke im Los "reproduziert", d. h. die Verteilungen von $X^L = X^L(N_L)$ und $X^S = X^S(n, N_L)$ sind bis auf die unterschiedliche Größe von N_L und n identisch.

Satz 6.1:

$$P(X^S(n, N_L) = x) = \int_0^1 b(x|\, n, p)\, dW(p) \qquad x = 0,1,\ldots, n$$

Beweis:

Es gilt für $X^L(N_L) = X$, $X^S(n, N_L) = x$ und $y := X - x$:

$$(6.1)\quad h(x|n,X,N_L)\, b(X|N_L,p) = b(x|n,p)\, b(y|N_L-n,p)$$

Folglich:

$$P(X^S(n,N_L)=x) = \sum_{X=0}^{N_L} h(x|n,X,N_L) \int_0^1 b(X|N_L,p)\, dW(p)$$

$$= \int_0^1 \sum_{y=0}^{N_L-n} b(x|n,p)\, b(y|N_L-n,p)\, dW(p)$$

$$= \int_0^1 b(x|n,p)\, dW(p)$$

□

Bemerkung 6.1:

Ist W auf einen Punkt p konzentriert, so ist die Anzahl der schlechten Stücke im Restlos $Y = Y(N_L - n)$ wegen (6.1) unabhängig von $X^S = X^S(n, N_L)$. Hieraus folgt, daß aus dem Stichprobenergebnis keine Rückschlüsse auf Losrest gezogen werden können. Stichprobenziehung ist in diesem Fall nicht informativ, vgl. Mood (1943), Case and Keats (1982). □

Die Reproduktion-Eigenschaft der Stichprobenverteilung erzwingt unter gewissen Zusatzvoraussetzungen, daß X^L und damit auch X^S gemischt-binomial-verteilt ist.

Es sei F_{N_L} die Verteilungsfunktion von $X^L(N_L)$ $N_L = 1,2,\ldots,$ (X^L nicht notwendig gemischt-binomial-verteilt!). Zu jedem N_L sei dann $W_{N_L}(p)$ $p \in (0,1)$ definiert durch:

$$W_{N_L}(p) := P\,(X^L(N_L) \leq Np)$$

Zu den W_{N_L} existiere ein Grenzwert W*, für den gilt:

(VI) $$W_{N_L} \xrightarrow[N_L \to \infty]{\text{schwach}} W^*$$

(VI) ist erfüllt, wenn:

(6.2) $$\lim_{N_L \to \infty} W_{N_L}(p) = W^*(p)$$

für alle Stetigkeitspunkte von W*(p) gilt.

Die Stichprobenverteilung ist von n und N_L abhängig und sei durch die Verteilungsfunktion G_{n,N_L} beschrieben. Die Reproduktionseigenschaft der Stichprobenverteilung lautet dann:

$$(V\ II) \qquad G_{n,\,N_L} = F_n \qquad \begin{matrix} n = 0, 1, \ldots, N_L \\ N_L = 1,2, \ldots \end{matrix}$$

Der angekündigte Satz lautet:

Satz 6.2:

Aus (VI) und (VII) folgt:

$$P(X^L(N_L) = X) = \int_0^1 b(X|N_L, p)\, dW^*(p)$$

$$X = 0, 1, \ldots, N_L \qquad \square$$

Einen Beweis dieses Satzes findet man bei Hald (1981), S. 318 ff.

Abschnitt 6.2:

CUSUM-Karten zum Aufdecken einer Prozeßverschlechterung

Ein Ziel der Stichprobenentnahme ist es, die Anzahl der verbliebenen schlechten Stücke $Y = Y(N_L - n)$ im Restlos zu schätzen. Dies geschieht meist über den Posteriori Erwartungswert $E(Y|X^S = x)$.

Dabei wird als Prozeßkurve häufig die zur Binomial-Verteilung konjugierte Beta-Verteilung gewählt. In diesem Fall ist die Berechnung von $E(Y|X^S = x)$ sehr einfach:

$$(6.3) \qquad E(Y|X^S = x) = (N_L - n)\ \frac{\alpha + x}{\alpha + \beta + n}$$

wobei α und β die Formparameter der Beta-Verteilung sind.

In vielen Fällen besteht die Annahme-/Ablehnung-Entscheidung über ein Los darin festzustellen, ob der geschätzte Prozentsatz schlechter Stücke im Restlos einen vorgegebenen Grenzwert überschreitet oder nicht:

$$(6.4) \qquad E\left(\frac{Y}{N_L - n} \mid X^S = x\right) = \frac{\alpha + x}{\alpha + \beta + n} \gtrless p_r$$

Hierbei ist p_r eine vorgegebene Trennqualität.

Bei kleinem Stichprobenumfang kann der Einfluß der Prozeßkurve auf die Annahme-/Ablehnungs-Entscheidung groß sein.

Dieses Beispiel zeigt, wie wichtig es ist, die Stichprobenergebnisse nicht nur zur Kontrolle des Ausschußanteils einzelner Lose zu verwenden, sondern auch zur Kontrolle der Prozeßkurve. Dies kann in einfacher Weise über CUSUM-Kontrollkarten geschehen.

Die Konsequenzen eines Alarms aufgrund einer CUSUM-Karte können verschiedener Art sein:

Bei der Warenausgangsprüfung:

- Stoppen der Produktion und Neueinrichten des Produktionsprozesses
- Anpassen der Annahme-/Ablehnungs-Entscheidung an die veränderte Prozeßkurve

Bei der Wareneingangsprüfung:

- Abbruch der Warenannahme
- Änderung der Kundenbewertung, Verschärfung der Prüfung

Im folgenden nehmen wir an, daß die Prozeßkurve durch die Dichte w(p) einer Beta-Verteilung gegeben ist mit:

$$w(p) = \frac{1}{Be(\alpha,\beta)} \; p^{\alpha-1} \; (1-p)^{\beta-1} \qquad p \in (0,1)$$
$$\alpha, \beta > 0$$

Es gilt:

$$E(P) = \bar{p} = \frac{\alpha}{\alpha+\beta}$$

$$VAR(P) = \sigma^2 = \frac{\bar{p}(1-\bar{p})}{\alpha+\beta+1}$$

Da im weiteren Text nur noch die Stichprobenergebnisse X^S und ihre Verteilung behandelt werden, können wir auf Index S zur Unterscheidung von den Losergebnissen im folgenden verzichten.

Die Verteilung von X: = X^S ist eine Polyaverteilung mit:

$$P(X=x) = \int_0^1 b(x|n,p)\,\frac{1}{Be(\alpha,\beta)}\,p^{\alpha-1}(1-p)^{\beta-1}\,dp$$

$$= \binom{n}{x}\,\frac{Be(\alpha+x,\beta+n-x)}{Be(\alpha,\beta)}$$

(6.5)
$$= \binom{n}{x}\,\frac{(\alpha+x-1)\cdots\alpha\,(\beta+n-x-1)\cdots\beta}{(\alpha+\beta+n-1)(\alpha+\beta+n-2)\cdots(\alpha+\beta)}$$

Es gilt:

$$E(X) = n\bar{p}$$
$$VAR(X) = n(n-1)\sigma^2 + n\bar{p}(1-\bar{p})$$

Abbildung 6.1 zeigt $f(x|n, \bar{p}, \sigma^2) := P(X = x)$ $(x = 0, 1, 2, \ldots)$ bei konstantem Erwartungswert $n\bar{p} = 125 * 0.65\ \% = 0.81$ und verschiedenen Werten σ für die Standardabweichung der Prozeßkurve.

Der Sollzustand der Prozeßkurve sei durch die Parameter α_S und β_S mit

$$\bar{p}_S = \frac{\alpha_S}{\alpha_S + \beta_S} \quad \text{und} \quad \sigma_S^2 = \frac{\bar{p}_S(1-\bar{p}_S)}{\alpha_S + \beta_S + 1}$$

bestimmt. Analog bestimmen α_A und β_A die Prozeßkurve im Alarmzustand.

Für den Likelihoodquotienten erhält man aus (6.5):

$$\frac{f(x|n,\alpha_A,\beta_A)}{f(x|n,\alpha_S,\beta_S)} =$$

(6.6)
$$\frac{(\alpha_S+\beta_S+n-1)\cdots(\alpha_S+\beta_S)(\alpha_A+x-1)\cdots\alpha_A\,(\beta_A+n-x-1)\cdots\beta_A}{(\alpha_A+\beta_A+n-1)\cdots(\alpha_A+\beta_A)(\alpha_S+x-1)\cdots\alpha_S\,(\beta_S+n-x-1)\cdots\beta_S}$$

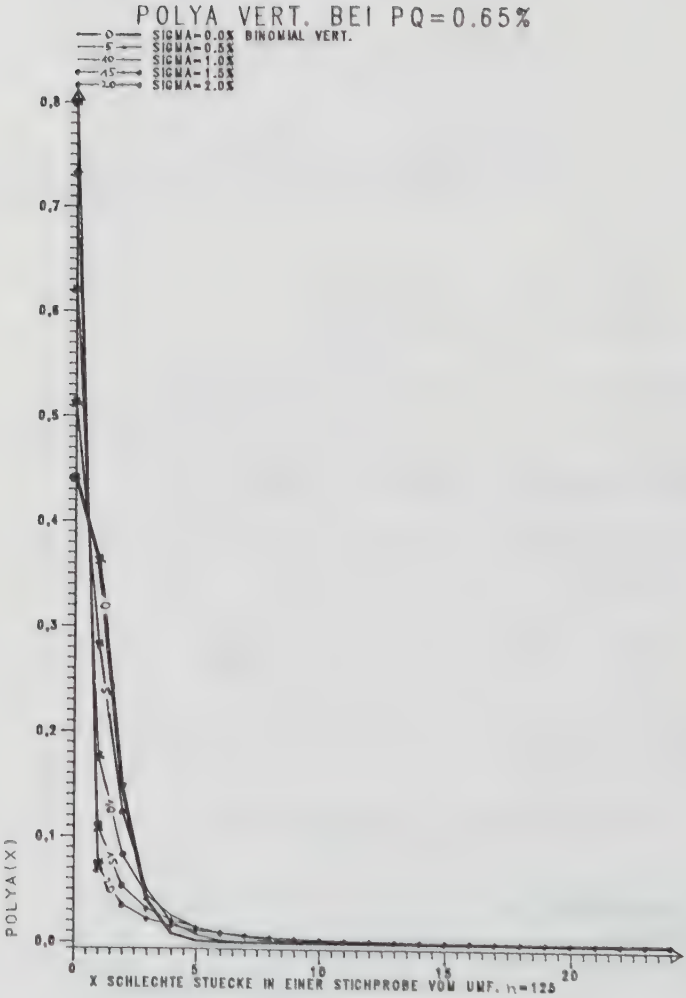

Abb. 6.1 : Dichte von Polya_Verteilungen mit gleichem Erwartungswert $n\bar{p}$ = 125*0.65% und verschiedenen Werten fuer σ .

Die Anzahl möglicher Alternativen α_A und β_A bzw. $\bar{p}_A$ und σ_A^2, die eine Verschlechterung des Produktionsprozesses repräsentieren, ist groß. Wir wollen im folgenden 3 Alternativen diskutieren.

Das einfachste Modell für eine verschlechterte Prozeßkurve ist:

(A_1) $\qquad \bar{p}_A > \bar{p}_S \qquad \sigma_A^2 = \sigma_S^2$

Die beiden folgenden Modelle benutzen eine inhaltliche Interpretation der Funktionalparameter α und β der Prozeßkurve. Wegen

$$E\left(\frac{Y}{N_L - n} \mid X = x\right) = \frac{\alpha + x}{(\alpha + x) + (\beta + n - x)}$$

ist die in der Prozeßkurve enthaltene Information über den Schlechtanteil im Restlos äquivalent mit einer zusätzlichen Stichprobe von Umfang $\alpha + \beta$ mit α schlechten und β guten Stücken. Die Beurteilung des Restloses erfolgt dann durch Zusammenfassen der beiden Stichproben mit $\alpha + x$ schlechten und $\beta + n - x$ guten Stücken.

Eine Prozeßverschlechterung läßt sich damit durch die α und β Parameter folgendermaßen modellieren:

Die Zusatzstichprobe enthält ein schlechtes Stück mehr, d. h.

(A_2) $\qquad \alpha_A = \alpha_S + 1 \qquad \beta_A = \beta_S$

Die Anzahl guter Stücke in der zusätzlichen Stichprobe hat sich verringert, d. h.:

(A3) $\qquad \alpha_A = \alpha_S \qquad \beta_A < \beta_S$

Im Fall (A_2) vereinfacht sich der Likelihoodquotient (6.6) beträchtlich. Wir erhalten für $\alpha_A = \alpha_S + 1$ und $\beta_A = \beta_S$:

$$\frac{f(x|n,\alpha_A,\beta_A)}{f(x|n,\alpha_S,\beta_S)} = \frac{\alpha_S+\beta_S}{\alpha_S+\beta_S+n} \quad \frac{\alpha_S+x}{\alpha_S}$$

$$= \frac{E_S(Y|X=x)}{E_S(Y)}$$

Hierbei ist $E_S(Y)$ der a-priori-Erwartungswert von Y unter der Prozeßkurve im Sollzustand und $E_S(Y|X = x)$ der entsprechende posteriori-Erwartungswert.

Die zu (A_2) gehörige CUSUM-Karte hat im Quotienten-Schema damit eine einfache, leicht interpretierbare Gestalt:

$$Q_0 = 1$$

$$Q_t = \max\left\{1, Q_{t-1} \frac{E_S(Y|X_t=x_t)}{E_S(Y)}\right\} \qquad t \geq 1 \tag{6.7}$$

Bevor wir auf die zu (A_3) zugehörige CUSUM-Karte eingehen, wollen wir kurz die Approximation der Beta-Binomial (=Polya)-Verteilung durch eine Gamma-Poisson (= negative Binomial)-Verteilung beschreiben.

Die Approximation beruht darauf, daß für kleine Werte von p die Binomial-Verteilung gut eine Poisson-Verteilung mit $\lambda = np$ approximiert wird. Ersetzt man die Beta-Verteilung durch die zur Poisson-Verteilung konjugierte Gamma-Verteilung mit der Dichte:

$$w(\lambda) = \frac{1}{\Gamma(\rho)} e^{-\tau\lambda} (\tau\lambda)^{\rho-1} \qquad \begin{matrix}\lambda > 0 \\ \rho, \tau > 0\end{matrix}$$

und

$$E(\lambda) = \frac{s}{\tau}$$

$$VAR(\lambda) = \frac{s}{\tau^2}$$

so erhält man für die Gamma-Poisson-Verteilung:

$$f(x|n,s,\tau) = \int_0^\infty \frac{e^{-(n+\tau)\lambda}(n\lambda)^x(\tau\lambda)^{s-1}}{\Gamma(s)\,x!}\,d\lambda$$

(6.7)

$$= \frac{\Gamma(s+x)}{x!\,\Gamma(s)}\,\eta^s(1-\eta)^x$$

wobei $\eta = \frac{\tau}{\tau+n}$

Für ganzzahlige Werte von s erhält man:

(6.8) $$f(x|n,s,\tau) = \binom{s-1+x}{s-1}\eta^s(1-\eta)^x$$

Für konstantes s und n bilden die Gamma-Poisson-Verteilungen eine 1-parametrige Exponentialfamilie mit

(6.9) $$\varphi = \ln(1-\eta) = \ln\frac{n}{n+\tau}$$

Für s = 1 erhält man eine geometrische Verteilung mit dem Parameter $\frac{n}{n+\tau}$.

Die Anpassung der Parameter s und τ an die Parameter α und β erfolgt aufgrund der inhaltlichen Interpretation von α und β durch:

(6.10) $$s = \alpha$$
$$\tau = \alpha+\beta$$

Mit dieser Kombination gilt:

(6.11) $$E(P) = \bar{p} = \frac{\alpha}{\alpha+\beta} = \frac{s}{\tau} = E(\lambda)$$

$$\frac{VAR(\lambda)}{VAR(P)} = \frac{\bar{p}\,(\alpha+\beta+1)}{(\alpha+\beta)\,\bar{p}\,(1-\bar{p})} \qquad (6.12)$$

$$= \frac{1 + \frac{1}{\alpha+\beta}}{1-\bar{p}}$$

Für kleine Werte von $\bar{p}$ und $\frac{1}{\alpha+\beta}$ können die beiden Prozeßkurven hinsichtlich Erwartungswert und Varianz als gut angepaßt gelten. Für $\alpha = s \gtrless 1$ sind beide Dichten glockenförmig bzw. monoton fallend.

In Tabelle 6.1 wird die Güte der Approximation der Polya-Verteilung durch eine Gamma-Poisson-Verteilung an einem Beispiel demonstriert. Für n = 125, $\bar{p} = \sigma = 0.65$ % und s = 1, τ = 153 werden für x = 0, 1, 2, ... der Wert der Dichte der Polya-Verteilung (POLYA(x)) und der entsprechende Wert der Gamma-Poisson-Verteilung (GAMPOIS(x)) berechnet. In der 4. Spalte ist der relative Fehler Δ_{rel} und in der 5. Spalte $\sum^{x} = \sum_{y=0}^{x}$ POLYA(y) angegeben. Mit Wahrscheinlichkeit 0.99 (0.999) ist der relative Fehler durch die Gamma-Poisson-Approximation kleiner als 1 % (5 %).

x	POLYA(X)	GAMPOIS(X)	Δ_{rel}	$\sum^{x}$
0	0.54960	0.55036	-0.00101	0.54980
1	0.24753	0.24746	0.00028	0.79734
2	0.11169	0.11127	0.00378	0.90903
3	0.05029	0.05003	0.00505	0.95931
4	0.02256	0.02250	0.00295	0.98188
5	0.01008	0.01012	-0.00304	0.99196
6	0.00449	0.00455	-0.01327	0.99645
7	0.00199	0.00205	-0.02804	0.99844
8	0.00088	0.00092	-0.04770	0.99932
9	0.00039	0.00041	-0.07263	0.99970
10	0.00017	0.00019	-0.10330	0.99987
11	0.00007	0.00008	-0.14025	0.99994
12	0.00003	0.00004	-0.18417	0.99998
13	0.00001	0.00002	-0.23583	0.99999
14	0.00001	0.00001	-0.29621	1.00000

Tabelle 6.1 : Approximation einer Polya-Verteilung mit n=125, $\bar{p}$=0.65% und σ=0.65% durch eine Gamma-Poisson-Verteilung mit n=125 , s=1 und τ = 153 .

Mit der Gamma-Poisson-Approximation läßt sich nun die unter (A_3) beschriebene Prozeßverschlechterung einfach behandeln.

Wegen (6.10) lautet die unter (A_3) beschriebene Prozeßverschlechterung in den s und τ Parametern:

$$(A_3) \qquad s_A = s_S \qquad \tau_A < \tau_S$$

(A_3) ist äquivalent damit, daß die Prozeßkurve im Soll- und Alarmzustand denselben Variationskoeffizienten

$$\frac{\sigma}{\bar{p}} = \frac{\sqrt{s}}{\tau} \cdot \frac{\tau}{s} = \frac{1}{\sqrt{s}}$$

haben.

Unter (A_3) bilden die Dichten der Gamma-Poisson-Approximation eine 1-parametrige Exponentialfamilie in $\vartheta = \frac{n}{n+\tau}$. Das zugehörige CUSUM-Schema kann damit nach Satz 2.2 auf die folgende Form transformiert werden:

$$K_o = 0$$
$$(6.13) \qquad K_t = \max \{0,\ K_{t-1} + X_t - k\} \qquad t \geq 1$$

Hierbei ist:

$$(6.14) \qquad k = \frac{-\rho \ln \frac{\tau_A (n+\tau_S)}{\tau_S (n+\tau_A)}}{\ln \frac{n+\tau_S}{n+\tau_A}}$$

Obwohl die unter (A_1) - (A_3) angegebenen Alternativen alle eine Verschlechterung des Produktionsprozesses beschreiben, zeigen die zugehörigen CUSUM-Karten ein sehr unterschiedliches Verhalten.

Dies sei an folgendem numerischen Beispiel belegt:

Parameter der Prozesskurve						
	$\bar{p}$ in %	σ in %	α	β	s	τ
Sollzustand	0.65	0.65	1.0	152	1.0	153
Alternative A1	2.14	0.65	10.8	495.7	10.8	506.5
Alternative A2	1.3	0.91	2.0	152	2.0	154
Alternative A3	2.14	2.14	1.0	45.7	1.0	46.7

Als Stichprobenumfang wird n = 125 gewählt.

Es bezeichne $f(x|S)$ bzw. $f(x|A_i)$ $(i = 1, 2, 3)$ die Dichte der Stichprobenverteilung in Sollzustand bzw. unter der Alternative A_i $(i = 1, 2, 3)$ an der Stelle $x = 0, 1, 2, \ldots$

Abbildung 6.2 zeigt das unterschiedliche Verhalten von $\ln\left(\frac{f(x|A_i)}{f(x|S)}\right)$ $x = 0, 1, 2, \ldots$ für $i = 1, 2, 3$. Für A_3 bildet $\ln\left(\frac{f(x|A_3)}{f(x|S)}\right)$ eine Gerade in x, die die x-Achse bei k = 1.46 schneidet. Wie man sich sofort überzeugt, gilt dieser Sachverhalt allgemein bei CUSUM-Karten der Form (6.12).

Auffällig ist die unterschiedliche Bewertung von Stichproben mit einer hohen Anzahl von schlechten Stücken.

Die Abbildungen (6.3 A_1) - (6.3 A_3) zeigen die ARL-Isolinien der CUSUM-Karten zu (A_1) - (A_3), wenn die wahre Stichprobenverteilung eine Gamma-Poisson-Verteilung mit n = 125 und $\bar{p}$ und σ ist. Hierbei variiert $\bar{p}$ von 0.4 % bis 1.8 % und σ von 0.0 % (Poisson-Verteilung) bis 2.2 %. Als Alarmgrenze wurde für alle 3 Karten $h^{CU} = 3.5$ gewählt.

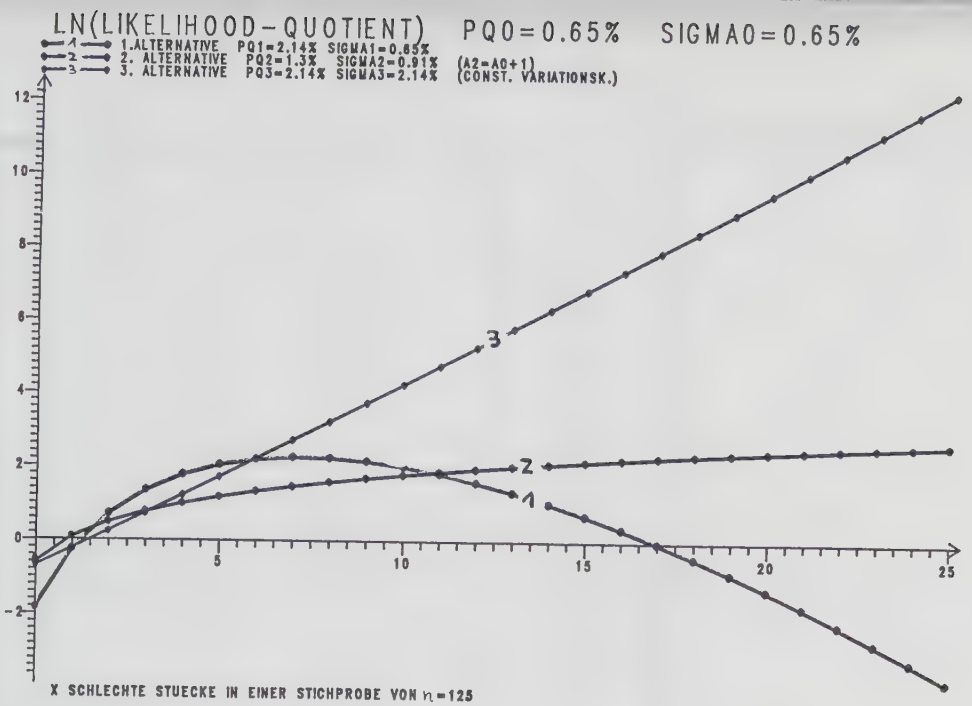

Abb. 6.2 : Das Verhalten der Zuwachsfunktion $Y(x)=\ln(f(x|A_i)/f(x|S))$ $(i=1,2,3)$ der zu den Alternativen A_i zugehoerigen CUSUM-Karten . Dichtefunktion : Polya mit $n=125$, $\bar{p}$ und σ .

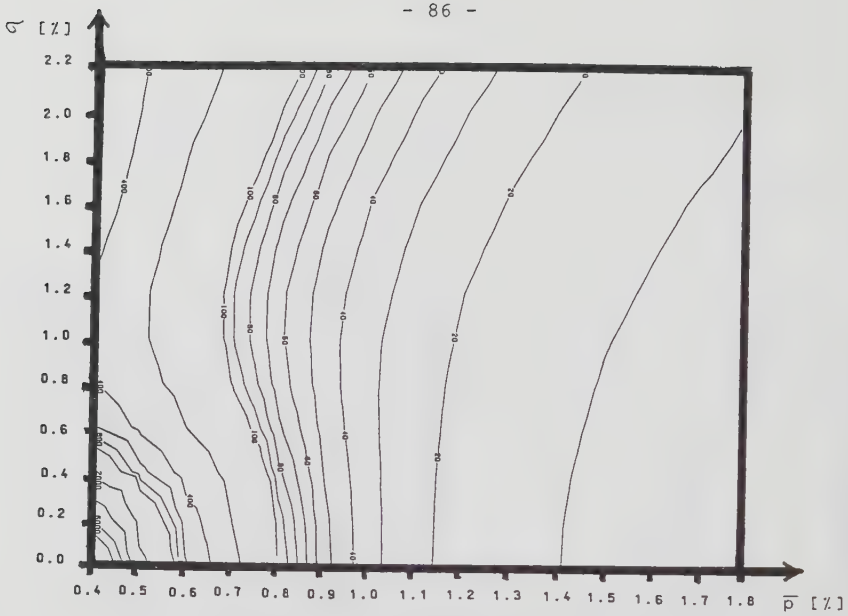

Abb. 6.3.A1 : ARL-Isolinien der CUSUM-Karte zur Alternative A1
Prozesskurve im Sollzustand : $\bar{p}=0.65\%$ $\sigma=0.65\%$
Prozesskurve im Altenativzustand : $\bar{p}=2.14\%$ $\sigma=0.65\%$
Wert der Alarmgrenze : $h^{CU}=3.5$

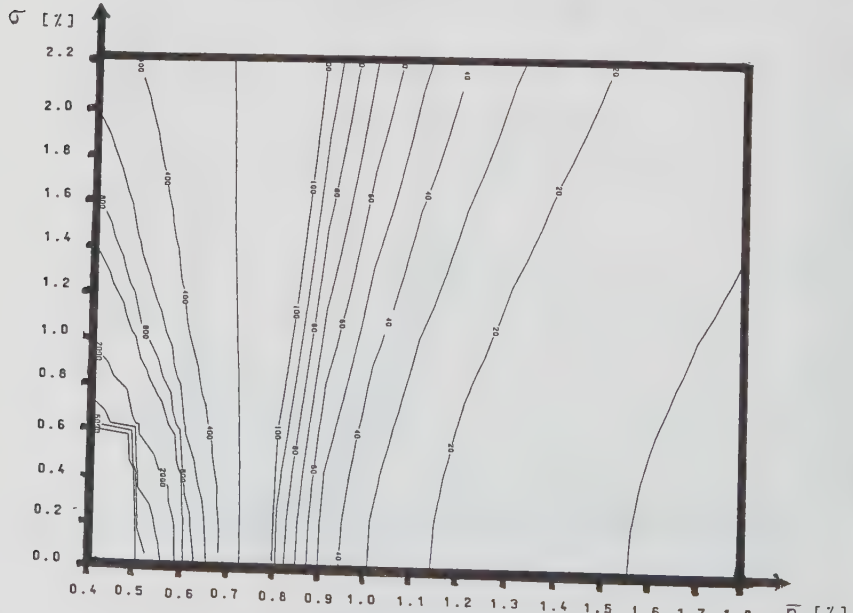

Abb. 6.3.A2 : ARL-Isolinien der CUSUM-Karte zur Alternative A2
Prozesskurve im Sollzustand : $\bar{p}=0.65\%$ $\sigma=0.65\%$
Prozesskurve im Altenativzustand : $\bar{p}=1.30\%$ $\sigma=0.91\%$
Wert der Alarmgrenze : $h^{CU}=3.5$

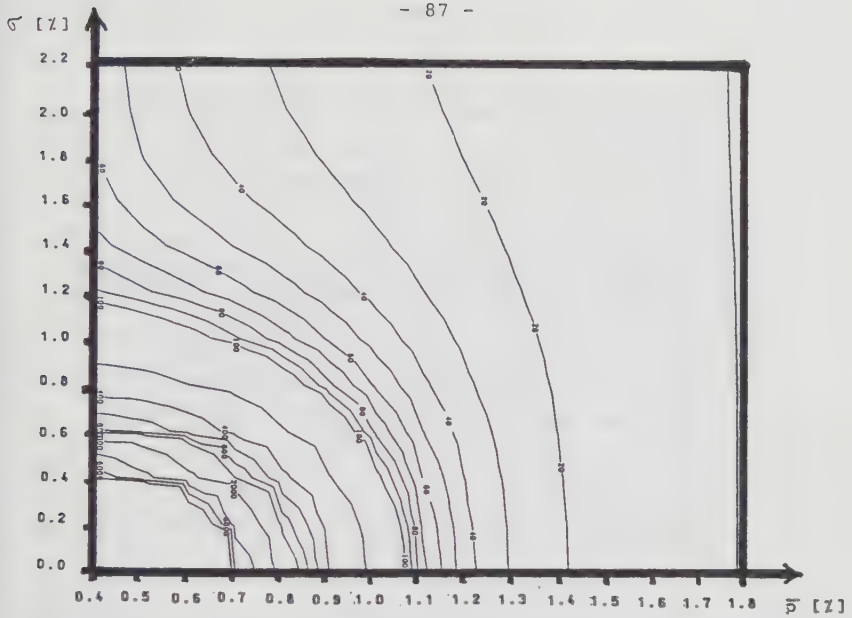

Abb. 6.3.A3 : ARL-Isolinien der CUSUM-Karte zur Alternative A3
Prozesskurve im Sollzustand : $\bar{p}$=0.65% σ=0.65%
Prozesskurve im Altenativzustand : $\bar{p}$=2.14% σ=2.14%
Wert der Alarmgrenze : h^{CU}=3.5

In allen 3 Fällen ist $ARL_i\ (\bar{p},\sigma)$ i = 1, 2, 3 monoton fallend in $\bar{p}$. Dies spiegelt die Tatsache wider, daß alle 3 CUSUM-Karten Verschlechterungen der Prozeßkurve erkennen sollen.

Der Einfluß der Prozeßvarianz σ^2 auf $ARL_i\ (\bar{p},\sigma)$ (i = 1, 2, 3) ist jedoch sehr unterschiedlich.

CUSUM-Karte zu (A_1)

Für $\bar{p} \leq 1.0$ % ist $ARL_1(\bar{p},\sigma)$ für festes $\bar{p}$ unimodal in σ.
Für $\bar{p} > 1.0$ % ist $ARL_1\ (\bar{p},\sigma)$ für festes $\bar{p}$ isoton in σ.

CUSUM-Karte zu (A_2):

Für $\bar{p} \leq 0.7$ % ist ARL_2 $(\bar{p},\sigma)$ für festes $\bar{p}$ monoton fallend, für $\bar{p} > 0.7$ monoton steigend in σ.

CUSUM-Karte zu (A_3):

$ARL_3(\bar{p},\sigma)$ ist für festes $\bar{p}$ eine monoton fallende Funktion in σ.

Der unterschiedliche Einfluß der Prozeßvarianz wird in der folgenden Tabelle 6.2 noch einmal verdeutlicht.

σ IN %	Alternative A1	Alternative A2	Alternative A3
0.0	224	228	5971
0.2	198	227	3198
0.4	152	224	1001
0.6	116	218	340
0.8	96	213	155
1.0	88	207	91
1.2	88	203	63
1.4	94	200	49
1.6	106	197	41
1.8	123	195	37
2.0	144	194	34
2.2	170	192	32

Werte der ARL von CUSUM-Karten zu versch. Alternativen

Tabelle 6.2 : Werte der ARL von CUSUM-Karten zu 3 verschiedenen Alternativen unter einer Prozesskurve mit $\bar{p}$ = 0.7% und Prozessstreuung σ

Das Verhalten der ARL-Isolinien kann folgendermaßen erklärt werden:

Bei n = 125 und X = o und X = 1 ist die Zuwachsfunktion Y(X) aller 3 CUSUM-Karten negativ oder Null (vgl. Abbildung 6.2). Die Abbildung 6.4 zeigt den Verlauf der Isolinien von P(X = 0) + P(X = 1) auf einem $(\bar{p},\sigma)$-Gitter. Für $\bar{p}$ = 0.6 % erhalten wir einen unimodalen Verlauf in Abhängigkeit von σ, vgl. Tabelle 6.3.

σ in %	Wert von P[X=0]+P[X=1] in %
0.0	82.6
0.2	82.3
0.4	81.7
0.6	81.5
0.8	81.9
1.0	82.8
1.2	83.9
1.4	85.1
1.6	86.3
1.8	87.4
2.0	88.5
2.2	89.4

Tabelle 6.3 : Werte von P[X=0]+P[X=1] fuer eine Polya-Verteilung mit n = 125, $\bar{p}$ = 0.6% und σ.

Für $\bar{p}$ > 0.7 % ist P(X = 0) + P(X = 1) isoton in σ.

Eine Vergrößerung von P(X = 0) + P(X = 1) wirkt in Richtung einer Verlängerung der Laufzeit der CUSUM-Karten. Dieser Effekt wird durch die unterschiedliche Bewertung von "Ausreißer"-Stichprobenergebnissen (siehe Abb. 6.2) unterschiedlich stark relativiert.

Vergleicht man die ARL-Isolinien aus Abbildung 6.3 A1 mit Tabelle 6.3 und Abbildung 6.4, so wird das unimodale Verhalten von ARL_1 $(\bar{p},\sigma)$ für festes $\bar{p}$ gut durch P(X = 0) + P(X = 1) beschrieben.

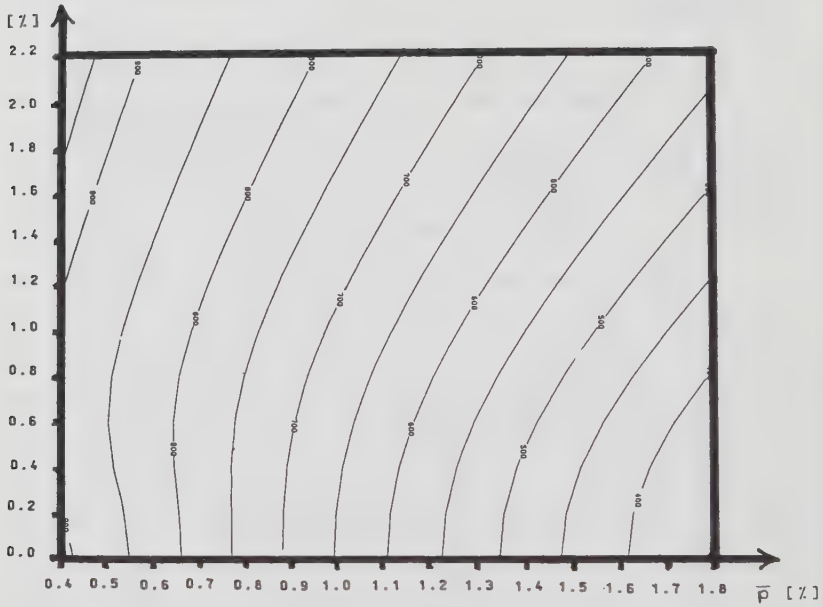

Abb. 6.4 : Isolinien von P(X=0)+P(X=1) fuer Gamma-Poisson Verteilungen mit n=125 und $\bar{p}$ und σ im oben angegeben Bereich .

Bemerkung 6.2:

Für sehr große Prozeßvarianzen σ^2 wird man auch für $ARL_i(\bar{p},\sigma)$ i=2, 3 für festes $\bar{p}$ ein unimodales Verhalten in σ erwarten dürfen. Abbildung 6.5 zeigt die Richtigkeit dieser Vermutung für $ARL_3(\bar{p},\sigma)$.

Insofern ist das unterschiedliche Verhalten von $ARL_i(\bar{p},\sigma)$ (i = 1, 2, 3) ein Resultat der Beschränkung auf den Bereich $\sigma \in [0.0\ \%, 2.2\ \%]$.

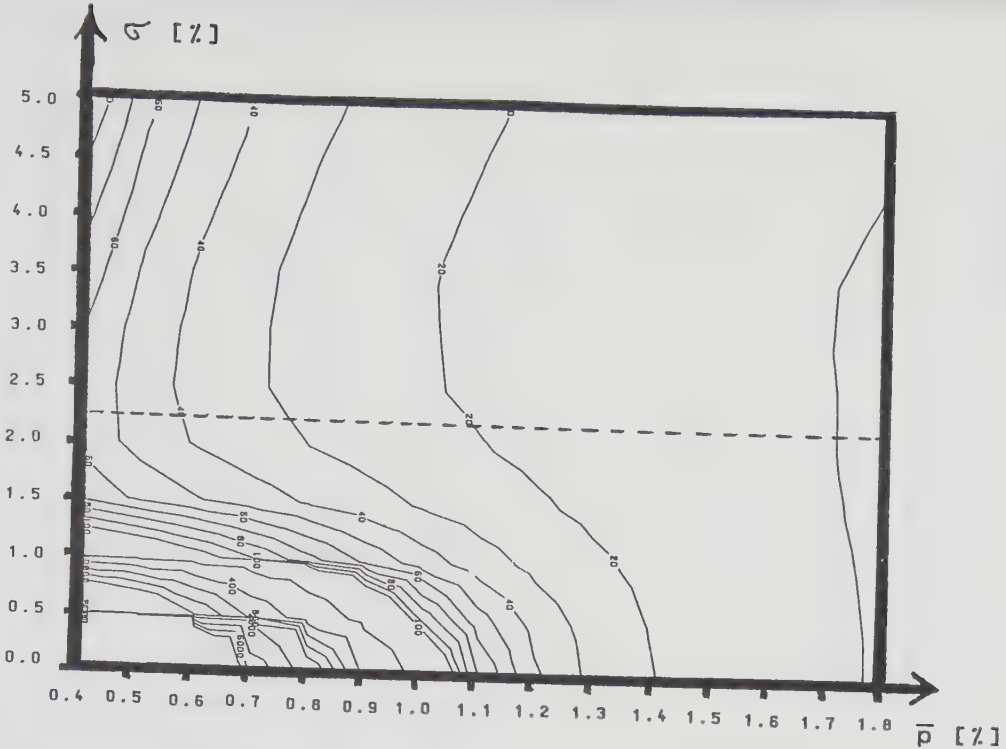

Abb. 6.5 : ARL-Isolinien der CUSUM-Karte zur Alternative A3 σ im Bereich [0.0%,5.0%] .
Gestrichelte Linie : σ =2.2%

Allerdings erscheint die Annahme wesentlich größerer Prozeßvarianzen nicht sehr realistisch. Für $N_L = 2000$ und $\bar{p} = 0.65$ % erhält man für den Erwartungswert der schlechten Stücke im Los $E(X^L(N_L)) = 13$ und für die Standardabweichung $\sqrt{VAR(X^L(N_L))}$:

σ in %	0.1	0.5	1.0	2.0	3.0
$\sqrt{VAR[X^L(N_L)]}$	4.1	10.6	20.3	40.1	60.0

□

Abschnitt 6.3:

CUSUM-Karten zum Erkennen einer Prozeßverbesserung

CUSUM-Karten können auch geführt werden, um Prozeßverbesserungen zu erkennen. In diesem Fall lauten die zu (A_1) - (A_3) analogen Alternativen zum Sollzustand:

(B_1) $\quad \bar{p}_B < \bar{p}_S \qquad \sigma_B = \sigma_S$

(B_2) $\quad \alpha_B = \frac{1}{2}\alpha_S \qquad \beta_B = \beta_S$

(B_3) $\quad \alpha_B = \alpha_S \qquad \beta_B > \beta_S$

Bemerkung 6.3:

Da häufig $\alpha_S \leq 1$ gilt, führt $\alpha_B = \alpha_S - 1 \leq 0$ auf nicht mehr zulässige Parameterwerte für α. Für $\alpha_S = 1$ entspricht (A_2) gerade einer Verdoppelung des α-Parameters. □

(B_1) - (B_3) werden für die folgenden Werte untersucht:

	Parameter der Prozesskurve					
	$\bar{p}$ in %	σ in %	α	β	s	τ
Sollzustand	0.65	0.65	1.0	152	1.0	153
Alternative B1	0.325	0.65	0.25	76.7	0.25	77
Alternative B2	0.325	0.46	0.5	152	2.0	152.5
Alternative B3	0.325	0.325	1.0	306	1.0	307

Die Abbildungen 6.6.B_1 - 6.6.B_3 zeigen die ARL-Isolinien für n = 125, 0.1 % $\leq \bar{p} \leq$ 1.0 % und 0.0 % $\leq \sigma \leq$ 2.2 % der 3 zugehörigen CUSUM-Karten. Als Alarmgrenze wurde $h^{CU} = 2.5$ für alle 3 Verfahren gewählt.

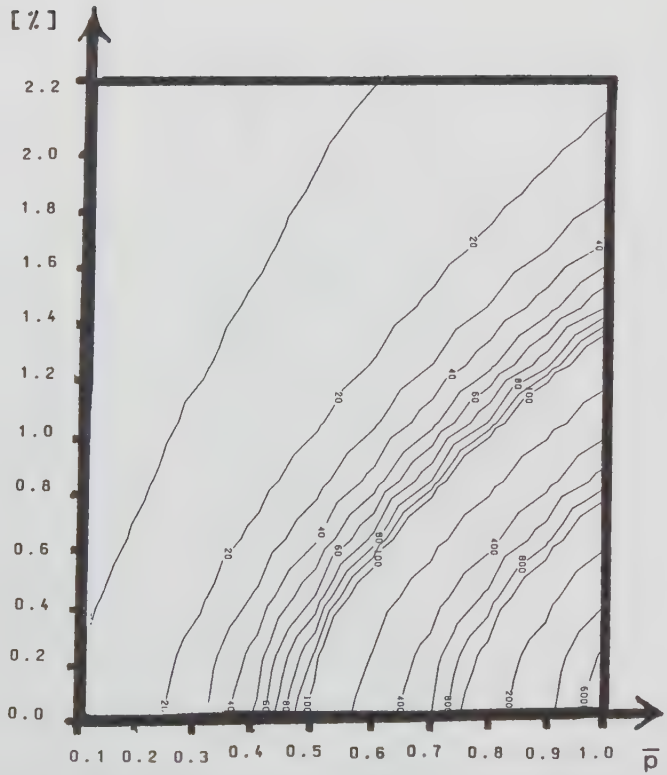

Abb. 6.6.B1 :
ARL-Isolinien der CUSUM-Karte zur Alternative B1
Prozesskurve im Sollzustand : $\bar{p}$=0.65% σ=0.65%
Prozesskurve im Alarmzustand : $\bar{p}$=0.325% σ=0.65%
Wert der Alarmgrenze : h^{CU}=2.5

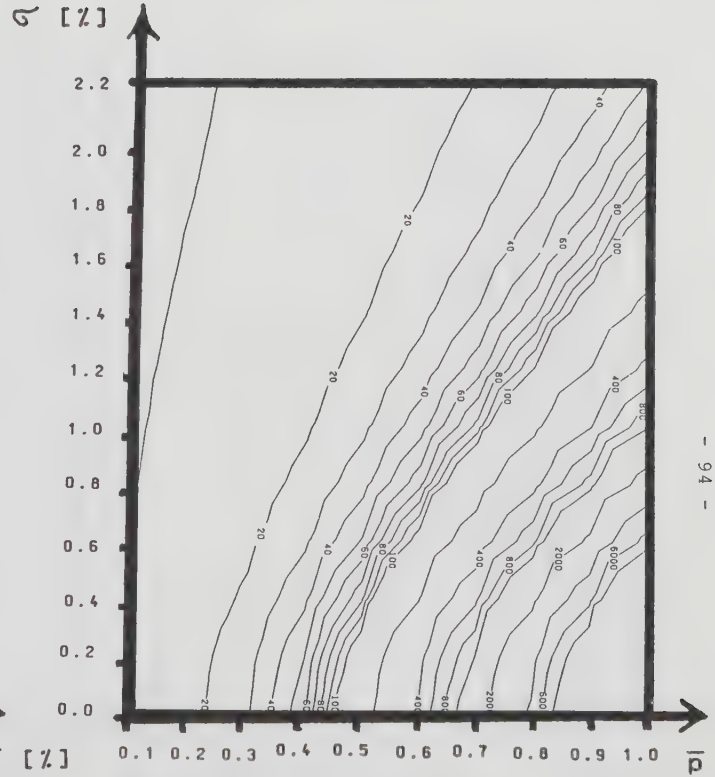

Abb. 6.6.B2 :
ARL-Isolinien der CUSUM-Karte zur Alternative B2
Prozesskurve im Sollzustand : $\bar{p}$=0.65% σ=0.65%
Prozesskurve im Alarmzustand : $\bar{p}$=0.325% σ=0.46%
Wert der Alarmgrenze : h^{CU}=2.5

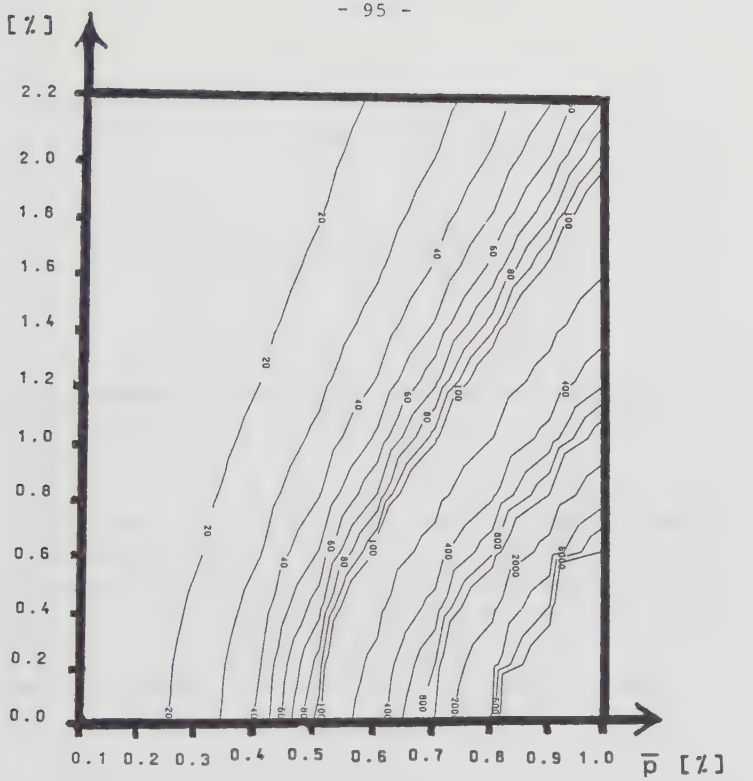

Abb. 6.6.B3 : ARL-Isolinien der CUSUM-Karte zur Alternative B3
Prozesskurve im Sollzustand : $\bar{p}$=0.65% σ=0.65%
Prozesskurve im Alarmzustand : $\bar{p}$=0.325% σ=0.325%
Wert der Alarmgrenze : h^{CU}=2.5

In allen 3 Fällen ist ARL_i $(\bar{p},\sigma)$ (i = 1, 2, 3) isoton in $\bar{p}$. Dies entspricht dem Sachverhalt, daß alle 3 Karten Prozeßverbesserungen erkennen sollen.

Der Einfluß der Prozeßvarianz ist anders als im vorhergehenden Abschnitt einheitlich: In allen Fällen ist ARL_i $(\bar{p},\sigma)$ für festes $\bar{p}$ monoton fallend in σ.

Dies hat seinen Grund darin, daß eine hohe Prozeßvarianz in der Stichprobenverteilung die Wahrscheinlichkeit für das Stichprobenergebnis X = 0 erhöht, vgl. Abb. 6.1.

Durch höhere Werte von h^{CU} läßt sich dieser Effekt mildern. Die Laufzeiten unter Prozeßkurven mit großen Werten von $\bar{p}$ und σ werden wesentlich erhöht, ohne daß die ARL für Prozeßkurven mit kleinen Werten von $\bar{p}$ und σ zu groß wird, vgl. Tabelle 6.4.

	Werte der ARL			
σ IN %	$\bar{p}$=0.3%		$\bar{p}$=0.7%	
	h^{CU}= 2.5	h^{CU}= 3.5	h^{CU}= 2.5	h^{CU}= 3.5
0.0	24.4	43.2	1024	27160
0.2	23.7	42.3	802	18546
0.4	21.9	39.6	470	7345
0.6	19.8	36.1	254	2528
0.8	18.8	32.4	147	978
1.0	16.6	29.3	94.3	459
1.2	15.5	26.8	66.5	256
1.4	14.6	24.7	50.6	163
1.6	13.8	23.2	40.8	115
1.8	13.2	21.9	34.2	87.6
2.0	12.8	20.7	29.6	69.8
2.2	12.4	19.8	26.2	57.8

Tabelle 6.4 : Der Effekt einer Vergroesserung von h^{CU} auf die ARL der CUSUM Karte zu Allernative B3 bei verschiedenen Werten von $\bar{p}$ und σ.

Kapitel 7

Adaptive Stichprobensysteme auf Grundlage von CUSUM-Karten

Im vorhergehenden Kapitel wurde untersucht, wie mit Hilfe einer CUSUM-Karte kontrolliert werden kann, ob die aktuelle Prozeßkurve von der bisher angenommenen Gestalt w_S abgewichen ist und eine alternative Gestalt w_A angenommen hat. In der Praxis liegen häufig mehrere Alternativen w_{A_i} $(i = 1, \ldots, N_A)$ für eine mögliche Abweichung der Prozeßkurve vor.

Diese Situation ist insbesondere dann gegeben, wenn die Kontrolle der Prozeßkurve unter dem Aspekt erfolgt, bei Änderung der Prozeßkurve den Prüfplan (n, c) an die neue Prozeßkurve anzupassen.

Eine Beschränkung auf endlich viele Alternativen w_{A_i} $(i = 1, \ldots, N_A)$ kann durch 2 Gründe motiviert sein:

(1) Aufgrund betrieblicher Erfahrungen ist bekannt, daß nur bestimmte Prozeßkurven während des Produktionsprozesses vorkommen.

(2) Man beschränkt sich auf eine Auswahl von Prüfplänen (n_i, c_i) $i = 0, 1, \ldots, N_A$ und damit implizit auf eine Auswahl zugehöriger Prozeßkurven w_{A_i} $(i = 0, 1, \ldots, N_A)$ hinsichtlich der die Prüfpläne (n_i, c_i) bayesoptimal sind. Hierbei ist $w_{A_0} = w_S$.

Für den Fall, daß endlich viele Alternativen w_{A_i} ($i = 1, \ldots, N_A$) als mögliche Abweichungen vom Sollzustand w_S betrachtet werden, liegt es nahe, die zugehörigen CUSUM-Karten simultan zu führen (vgl. Kapitel 5). Wenn dann (mindestens) eine CUSUM-Karte ihre Alarmgrenze $h^{CU}(i)$ ($i = 1, \ldots, N_A$) überschreitet, muß eine neue Prozeßkurve geschätzt werden.

Für die Schätzung $\hat{w}(p)$ der neuen Prozeßkurve erscheinen 2 Annahmen plausibel:

(SI) In die Schätzung $\hat{w}(p)$ sollen nur diejenigen Stichprobenergebnisse $X_t, X_{t-1}, \ldots, X_{t-L_t+1}$ eingehen, die zur Ablehnung der bisherigen Prozeßkurve $w_S(p)$ geführt haben. □

(SII) Es genügt, sich bei der Schätzung auf die Alternativen w_{A_i} ($i = 1, \ldots, N_A$) zu beschränken. □

Es bezeichne für $i = 1, \ldots, N_A$:

$$(7.1)\quad q_t(i)(\ell) := \begin{cases} \prod_{t'=t-\ell+1}^{t} \frac{f(X_{t'}|w_{A_i})}{f(X_{t'}|w_S)} & \ell \in \{1,2,\ldots,t\} \\ 1 & \ell = 0 \end{cases}$$

Satz 2.1 besagt dann gerade:

$$(7.2)\quad Q_t(i) = \max_{0 \le \ell \le t} q_t(i)(\ell) \qquad i = 1, \ldots, N_A,\ t \ge 0$$

Es bezeichne $L_t(i)$ ($i = 1, \ldots, N_A$) denjenigen Wert von ℓ, der zum Zeitpunkt t die rechte Seite von (7.2) maximiert. $L_t(i)$ ist gerade die Anzahl zurückliegender Daten, die in die Berechnung der i. CUSUM-Karte eingehen ($i = 1, \ldots, N_A$).

Es sei $i_{max}(t)$ derjenige Index $i \in \{1, \ldots, N_A\}$ dessen zugehörige CUSUM-Karte $Q_t(i_{max}(t))$ zum Zeitpunkt t den höchsten Wert hat:

$$(7.3) \quad Q_t(i_{max}(t)) = \max_{i \in \{1,\ldots,N_A\}} Q_t(i)$$

Satz 7.1:

$$(7.4) \quad Q_t(i_{max}(t)) = \max_{0 \le l \le t} \left(\max_{i \in \{1,\ldots,N_A\}} q_t(i)(l) \right)$$

Beweis:

Da die Maximumbildung in (7.4) jeweils über eine endliche Menge positiver Zahlen erfolgt, gilt:

$$\max_{0 \le l \le t} \left(\max_{i=1,\ldots,N_A} q_t(i)(l) \right) = \max_{i=1,\ldots,N_A} \left(\max_{0 \le l \le t} q_t(i)(l) \right)$$

$$= \max_{i=1,\ldots,N_A} Q_t(i)$$

$$= Q_t(i_{max}(t)) \qquad \square$$

Für die Schätzung einer neuen Prozeßkurve erscheint damit die Wahl

$$(7.5) \quad \hat{w} = w_{A_{i_{max}(t)}}$$

plausibel und mit dem CUSUM-Prinzip in einfacher Weise verträglich.

In die Schätzung unter (7.5) gehen $L_t(i_{max}(t))$ zurückliegende Beobachtungen ein. Wie aus (7.4) sofort folgt,

ist $\hat{w}$ eine ML-Schätzung bzgl. der Alternativen w_{A_i} ($i = 1, \ldots, N_A$) auf Basis der letzten $L_t(i_{max}(t))$ Beobachtungen. Damit liefern die CUSUM-Karten $Q_t(i)$ bzw. $CU_t(i)$ ($i = 1, \ldots, N_A$) nicht nur eine Entscheidung über den Zeitpunkt, wann eine Adaption an eine neue Prozeßkurve erfolgen soll, sondern auch gleichzeitig eine einfache Schätzmethode, die den Anforderungen (SI) und (SII) genügt.

Abbildung 7.1 dokumentiert das oben beschriebene Adaptionsverfahren in einem Pseudo-Code.

Adaptionsverfahren MULT.CUSUM in einen Pseudocode

```
PROGRAM  MULT.CUSUM
READ     Parameter :
         Prozesskurvenparameter im Sollzustand : p̄_S , σ_S
         Anzahl der alternativen Prozesskurven : N_A
         Parameter der Prozesskurvenaltenativen: p̄_i , σ_i  i=1,...,N_A
         Alarmgrenzen der CUSUM-Karten         : h^CU(i) i=1,...,N_A
BEGIN    (" Initialisierung des Verfahrens ")
         Berechnung des zu (p̄_S,σ_S) optimalen Pruefplans (n_S,c_S)
         Bestimmung der Zuwachsfunktionen Y[i](x) mit :
                  Y[i](x)=ln(f(x I n_S,p̄_Ai,σ_Ai)/f(x I n_S,p̄_S,σ_S))
                          x=0,1,...,n    i=1,...,N_A
         Setzen aller CUSUM-Karten auf 0 :  CU(i)=0   i=1,...,N_A
WHILE ( t≤ NZEIT )  DO                (" Schleife ueber NZEIT Lose ")
         Ziehen einer Stichprobe vom Umfang n_S
         Annahmeentscheidung : x(t)≤ c_S ?
         Aufdatieren der CUSUM-Karten : CU(i)=max{0,CU(i)+Y[i](x(t))}
                                                 i=1,...,N_A
         IF ( CU(i) ≥ h^CU(i)  fuer ein i∈{1,...,N_A} ) THEN
              Bestimme CUSUM-Karte CU(imax)=max CU(i)
              Neuer Sollzustand : p̄_S = p̄_imax  , σ_S = σ_imax
              Reinitialisierung des Verfahrens
         ENDIF
         Erhoehung des Losindex t um 1
ENDWHILE (" Ende der Lospruefung ")
END
```

Abb. 7.1 : Pseudocode eines adaptiven Stichprobensystems

Wir wollen die Eigenschaften des Adaptionsverfahrens an einem Beispiel diskutieren.

Wir nehmen an, daß die Prüf-, Garantie- und Reparaturkosten für ein Warenlos linear in den Einzelstücken sind und daß bei der Stichprobe gefundene schlechte Stücke durch gute Stücke ersetzt werden (Rectifying Inspection). Bei gegebenem Losumfang N_L und bekannter Trennqualität

$$p_r = \frac{\text{Prüfkosten pro Stück}}{\text{Garantiekosten pro Stück - Reparaturkosten pro Stück}} \tag{7.6}$$

minimiert ein bayesoptimaler Prüfplan (n*, c*) den erwarteten vermeidbaren Verlust (Regret), vgl. Hald (1981), Seite 29:

$$\begin{aligned} R((n,c)|w) = \min_{(n,c)} \Big\{ & n \int_0^{p_r} (p_r - p)\, w(p)\, dp + \\ & (N_L - n) \int_0^{p_r} (p_r - p)(1 - L_{n,c}(p))\, w(p)\, dp + \\ & (N_L - n) \int_{p_r}^{1} (p - p_r)\, L_{n,c}(p)\, w(p)\, dp \Big\} \end{aligned} \tag{7.7}$$

wobei $L_{n,c}(p) = \sum_{x=0}^{n} b(x|n,p)$ die Annahmewahrscheinlichkeit eines Loses unter Benutzung des Prüfplans (n, c) ist.

Der 1. Term der rechten Seite von (7.7) ist der durch die Stichprobenziehung verursachte vermeidbare Verlust, während die beiden übrigen Terme die erwarteten Verluste durch irrtümliche Ablehnung oder Annahme eines Loses beschreiben.

Der relative Verlust, der entsteht, wenn bei Vorliegen einer Prozeßkurve w der nicht-optimale Plan (n, c) benutzt wird, ist gegeben durch:

$$(7.8)\quad \ell_{rel}((n,c)|w) := \frac{R((n,c)|w) - R((n^*,c^*)|w)}{R((n^*,c^*)|w)}$$

Hierbei ist (n*, c*) der zur Prozeßkurve w bayesoptimale Prüfplan.

Abbildung 7.2 zeigt die Isolinien von $100 * R((n^*, c^*)|w)$, wobei $N_L = 2000$ und $p_r = 1.8 * 0.65\ \% = 1.17\ \%$ ist. Als Prozeßkurven werden hierbei Beta-Verteilungen angenommen. Ihr Erwartungswert $\bar{p}$ liegt im Bereich $[0.2\ \%, 1.8\ \%]$. Für die Prozeßvarianz σ^2 wurde der Bereich $\sigma \in [0.2\ \%, 2.2\ \%]$ gewählt. Für noch kleinere Varianzen ist die bayesoptimale Entscheidung entartet: Man erhält für $\bar{p} < p_r$ "Annahme des Loses ohne Stichprobenziehung" und für $\bar{p} > p_r$ "Ablehnung des Loses ohne Stichprobenziehung". Die Betrachtung noch größerer Prozeßvarianzen erscheint nach Bemerkung 6.2 nicht mehr realistisch.

Bemerkung 7.1:

Der Wert von $1.8 * 0.65\ \%$ für die Trennqualität p_r erklärt sich dadurch, daß AQL-Wert (vgl. Kapitel 8) und Trennqualität p_r durch $p_r \approx 1.8 * AQL$ in sinnvoller Weise miteinander gekoppelt werden können (vgl. Lenz und Rendtel (1984)).

□

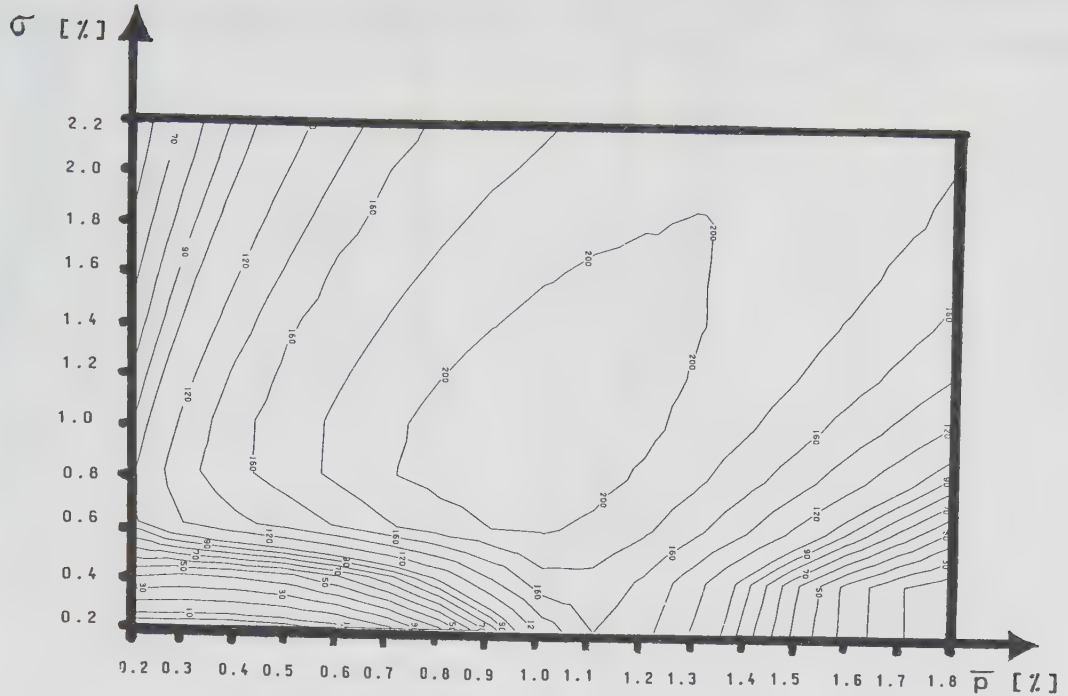

Abb. 7.2 : Isolinien von 100*R([n^{ψ},c^{ψ}] I w) . N_L=2000 , Trennqualitaet p_r=1.17% , Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ .

Die Abbildungen 7.3 a - 7.3 d zeigen die Isolinien von $\ell_{rel}((n_i, c_i)|w(\bar{p},\sigma))$, wobei $(n_1, c_1) = (125,2)$, $(n_2, c_2) = (50,1)$, $(n_3, c_3) = (125,1)$ und $(n_4, c_4) = (313,3)$. Die Vereinigung der 10 %-Bereiche von $\ell_{rel}((n_i, c_i)|w(\bar{p},\sigma))$ $i = 1, 2, 3, 4$ deckt große Bereiche des Parameterraums $(\bar{p},\sigma)$ ab, d. h. für die meisten der Prozeßkurven mit $(\bar{p},\sigma)$ ist einer der 4 Pläne (n_i, c_i) $i = 1, \ldots, 4$ fast bayesoptimal.

Lediglich in Bereichen, wo $R((n^*, c^*)|w(\bar{p},\sigma))$ klein ist, ist die Effizienz der Pläne (n_i, c_i) $i = 1, \ldots, 4$ gering, jedoch ist dort der absolute Verlust $R((n, c)|w) - R((n^*, c^*)|w)$ gering, vgl. Abb. 7.4 a - 7.4 d.

Übrig bleiben lediglich Bereiche, wo

- Annahme ohne Stichprobenprüfung

  oder

- Ablehnung ohne Stichprobenprüfung

bayesoptimal sind.

Im ersten Fall sind die mit der Stichprobenentnahme verbundenen Kosten der Preis, den man für Informationsbeschaffung über mögliche Abweichungen der Prozeßkurve zu bezahlen hat.

Die Prozeßkurven, die zur Ablehnung ohne Stichprobenprüfung führen, erscheinen unter praktischen Gesichtspunkten nicht mehr realistisch.

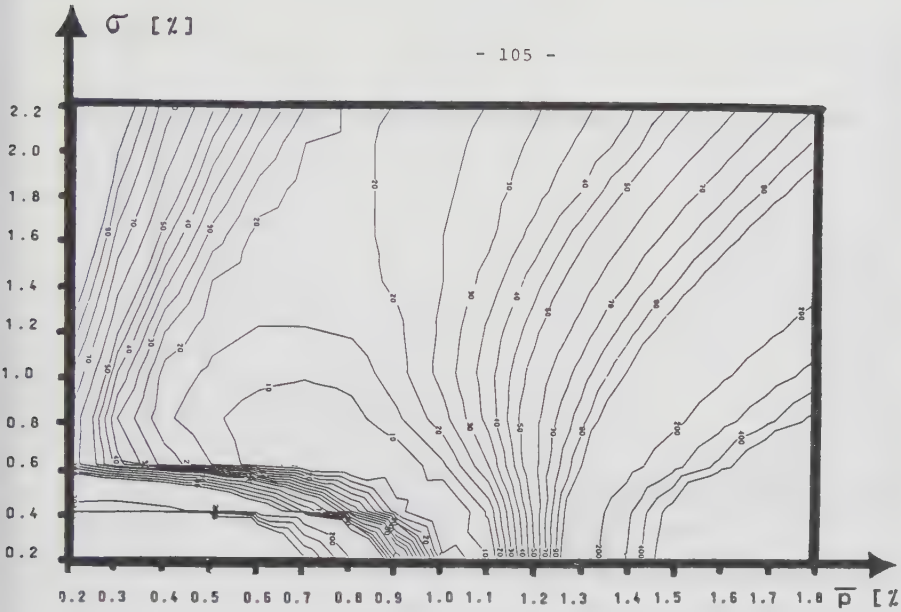

Abb. 7.3.a : Isolinien von 100*1vc([125,2] I w) , N_L=2000 , Trennqualitaet p_r=1.17% , Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ .

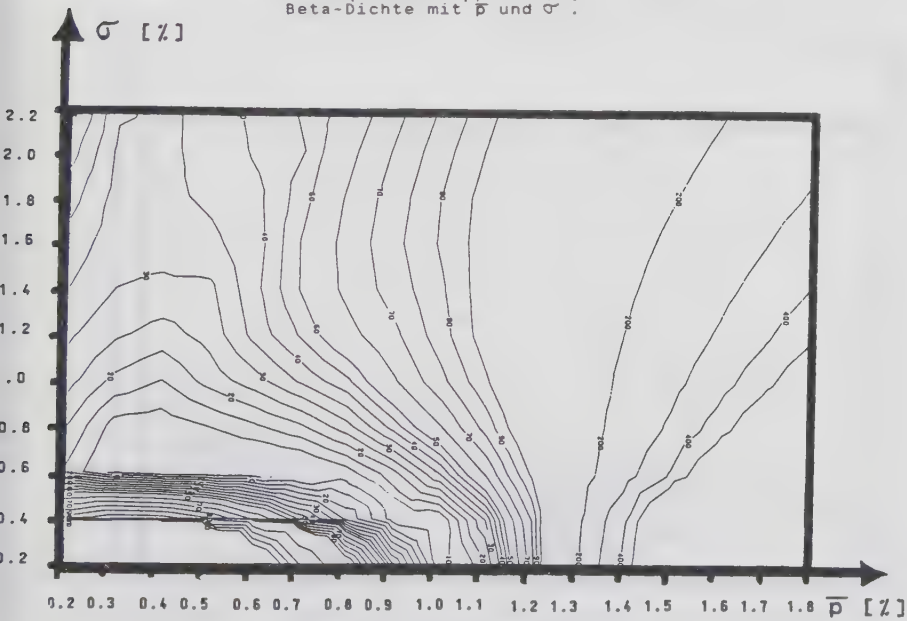

Abb. 7.3.b : Isolinien von 100*1vc([50,1] I w) , N_L=2000 Trennqualitaet p_r=1.17% , Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ .

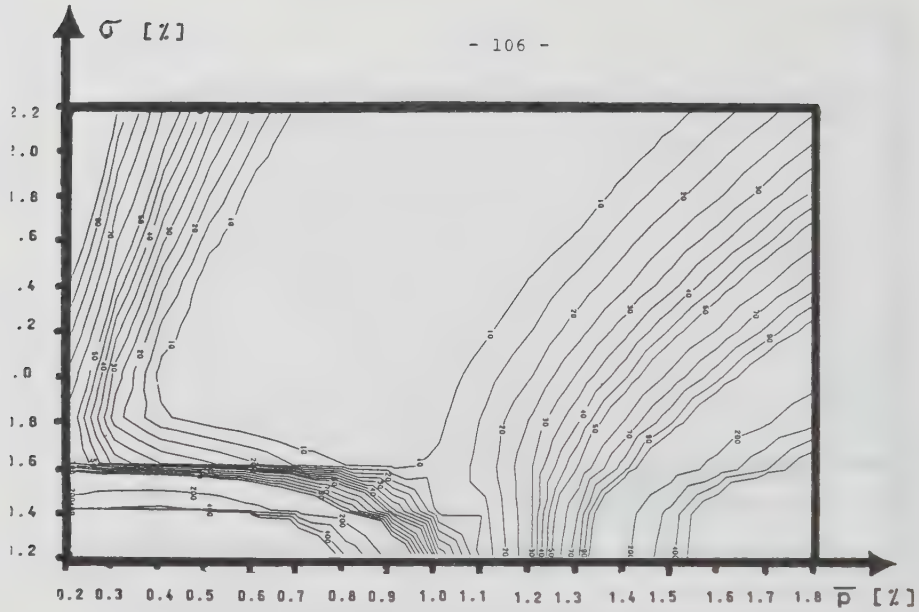

Abb. 7.3.c : Isolinien von 100*lvec([125,1] I w) , N_L=2000 , Trennqualitaet p_v=1.17% , Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ .

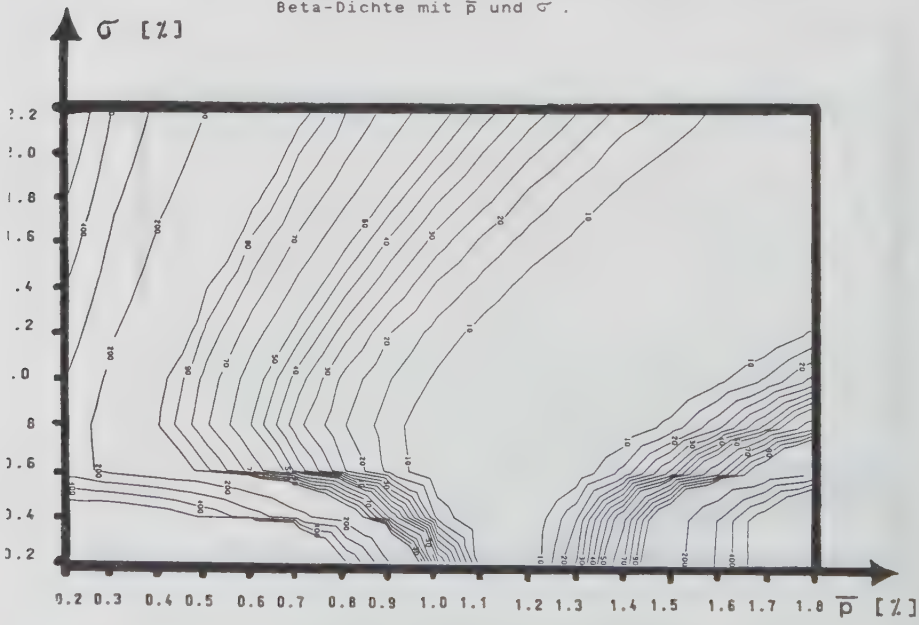

Abb. 7.3.d : Isolinien von 100*lvec([313,3] I w) , N_L=2000 , Trennqualitaet p_v=1.17% , Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ .

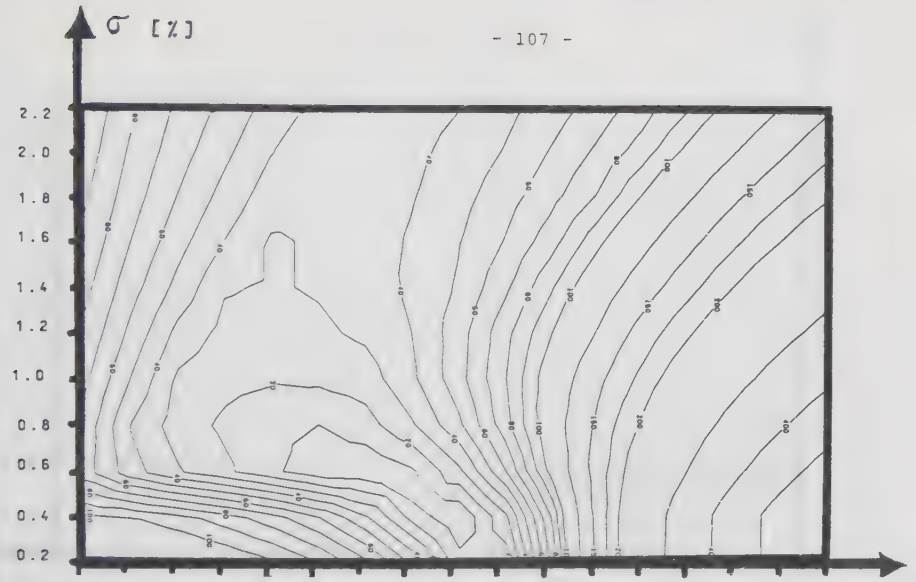

Abb. 7.4.a : Isolinien von 100*(R([125,2]I w)-R([n*,c*]I w))
N = 2000 , Trennqualitaet pv=1.17% ,
Prozesskurve : Beta-Dichte mit p̄ und σ .

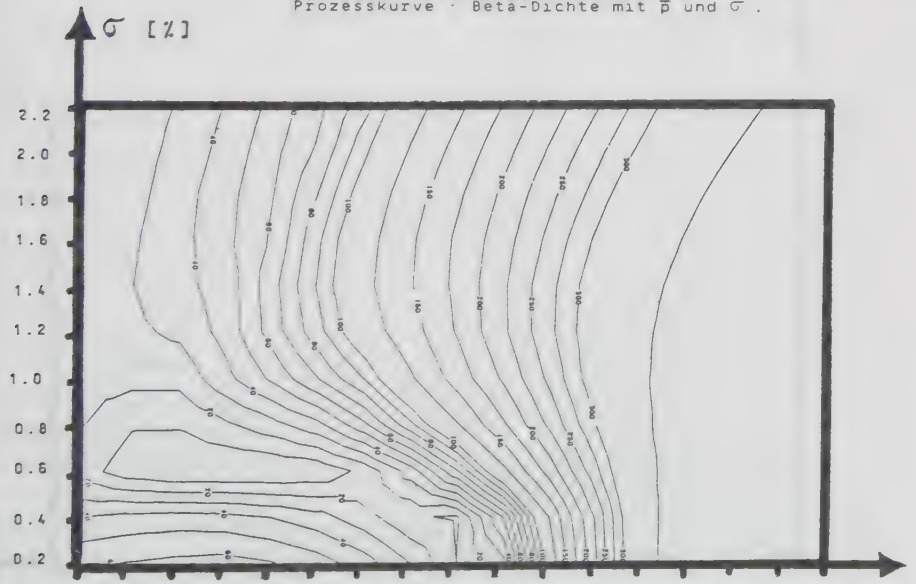

Abb. 7.4.b : Isolinien von 100*(R([50,1]I w)-R([n*,c*]I w))
N = 2000 , Trennqualitaet pv=1.17% ,
Prozesskurve : Beta-Dichte mit p̄ und σ .

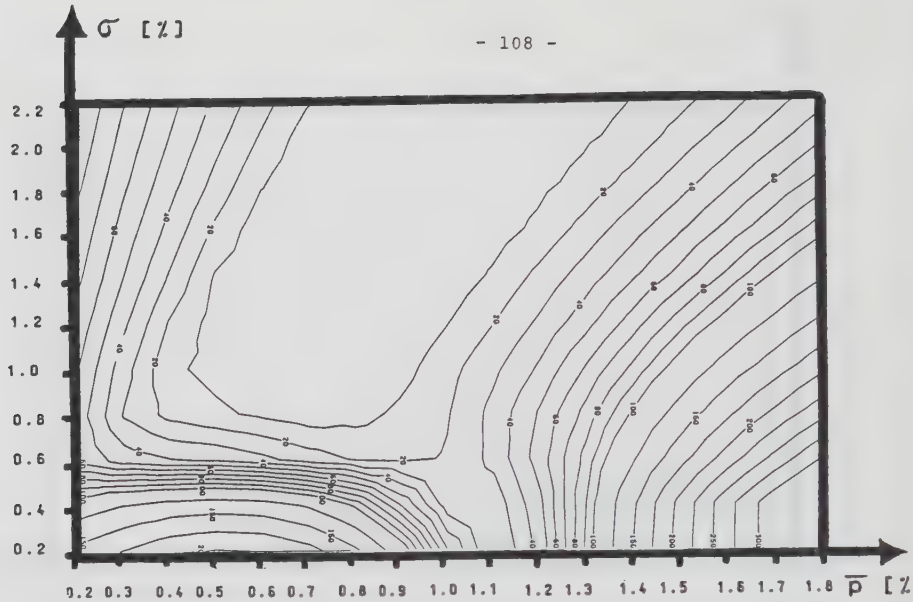

Abb. 7.4.c : Isolinien von 100*(R([125,1]I w)-R([n*,c*]I w))
N_L= 2000 , Trennqualitaet p_Y=1.17% ,
Prozesskurve : Beta-Dichte mit $\bar{p}$ und σ .

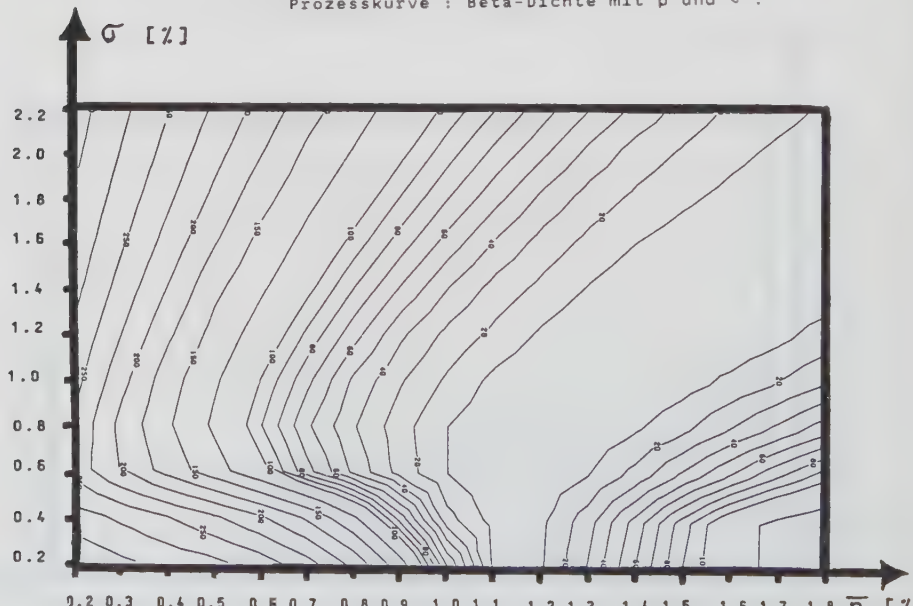

Abb. 7.4.d : Isolinien von 100*(R([313,3]I w)-R([n*,c*]I w))
N_L= 2000 , Trennqualitaet p_Y=1.17% ,
Prozesskurve : Beta-Dichte mit $\bar{p}$ und σ .

Wir wollen jetzt den Prüfplänen (n_i, c_i) $(i = 1, \ldots, 4)$ jeweils eine Prozeßkurve zuordnen, für die der zugehörige Prüfplan bayesoptimal ist. Für $N_L = 2000$ und $p_r = 1.17$ % erhält man (vgl. Lenz und Rendtel (1984)):

Die Zuordnung einer beta-verteilten Prozeßkurve mit $\bar{p}_S$ und σ_S zu dem Plan (125,2) ist eindeutig. Man erhält: $\bar{p}_S = 0.9$ %, $\sigma_S = 0.47$ %.

Für die anderen Pläne existiert im allgemeinen eine ganze Schar von $(\bar{p}, \sigma)$-Werten, so daß (n, c) bezüglich $w(\bar{p}, \sigma)$ bayesoptimal ist.

Die Wahl von w_{A_1} mit $\bar{p}_{A_1} = 0.4$ % und $\sigma_{A_1} = 0.6$ % erfolgte unter dem Kriterium, $\bar{p}_{A_1}$ unter allen Scharparametern zu (50,1) möglichst klein zu wählen.

Die Wahl von w_{A_2} mit $\bar{p}_{A_2} = 1.0$ % und $\sigma_{A_2} = 2.2$ % maximiert σ unter den gegebenen Scharparametern für (n, c) = (125,1).

w_{A_3} mit $\bar{p}_{A_3} = 1.3$ % und $\sigma = 1.0$ % minimiert $\bar{p}_{A_3}$ unter den Scharparametern für (n, c) = (313,3). Diese Wahl erfolgte unter der Annahme, daß Prozeßkurven mit einem Erwartungswert wesentlich oberhalb der Trennqualität nicht realistisch sind.

Wir wollen annehmen, daß das adaptive Stichprobensystem mit dem Plan (125,2) startet. Wir nehmen damit implizit an, daß die Prozeßkurve w_S mit $\bar{p}_S = 0.9$ % und $\sigma_S = 0.47$ % dem Produktionsprozeß zugrunde liegt. Wir führen dann die 3 CUSUM-Karten $CU_t(i)$ zu den Alternativen w_{A_i} $(i = 1, 2, 3)$. Falls eine der 3 Karten Alarm gibt, wählen wir diejenige Alternative $A_{i_{max}}(t)$ zum neuen Sollzustand w_S, deren CUSUM-Karte den höchsten Wert hat.

Wir haben noch kein Gütekriterium hinsichtlich der Wahl von $h^{CU}(1)$, $h^{CU}(2)$, $h^{CU}(3)$ angegeben.

Eine Möglichkeit besteht darin, die $h^{CU}(i)$ $i = 1, 2, 3$ so zu wählen, daß die ARL's der einzelnen CUSUM-Karten unter w_S einen vorgegebenen Wert haben.

Wir können aber auch die ARL unter einer Prozeßkurve w und die Verluste, die bei einem Nichterkennen von w auftreten, miteinander in Beziehung setzen. Die ARL des Gesamtsystems sollte kurz sein unter einer Prozeßkurve w, wo einer der alternativen Prüfpläne (n_{A_i}, c_{A_i}) $i = 1, 2, 3$ einen deutlich geringeren Verlust $R((n_{A_i}, c_{A_i})|w)$ hat als $R((n_S, c_S)|w)$, d. h. wo

$$\ell_1(w) := R((n_S, c_S)|w) - \min_{i=1,2,3} R((n_{A_i}, c_{A_i})|w) \tag{7.9}$$

groß ist, sollte die ARL des Gesamtsystems klein sein.

Abbildung 7.5 zeigt die Isolinien von $\ell_1(w(\bar{p},\sigma))$ in Abhängigkeit von $\bar{p}$ und σ.

Eine zweite Überlegung betrifft die Wahl des Prüfplans bei einer Prüfplanänderung. Es bezeichne $P_w(i)$ $(i = 1,2,3)$ die Wahrscheinlichkeit, daß das Gesamtsystem unter einer Prozeßkurve w die i. Alternative wählt $(i = 1, 2, 3)$, d. h. daß im Alarmzeitpunkt die i. CUSUM-Karte den höchsten Wert hat.

Der erwartete Verlust nach einer Prüfplanänderung ist unter einer Prozeßkurve w gegeben durch

$$\sum_{i=1}^{3} R((n_{A_i}, c_{A_i})|w)\, P_w(i)$$

Von diesem erwarteten Verlust muß jedoch noch der unvermeidliche Verlust abgezogen werden, der durch die Beschränkung auf die 3 Prüfpläne (n_{A_i}, c_{A_i}) $i = 1, 2, 3$ gegeben ist.

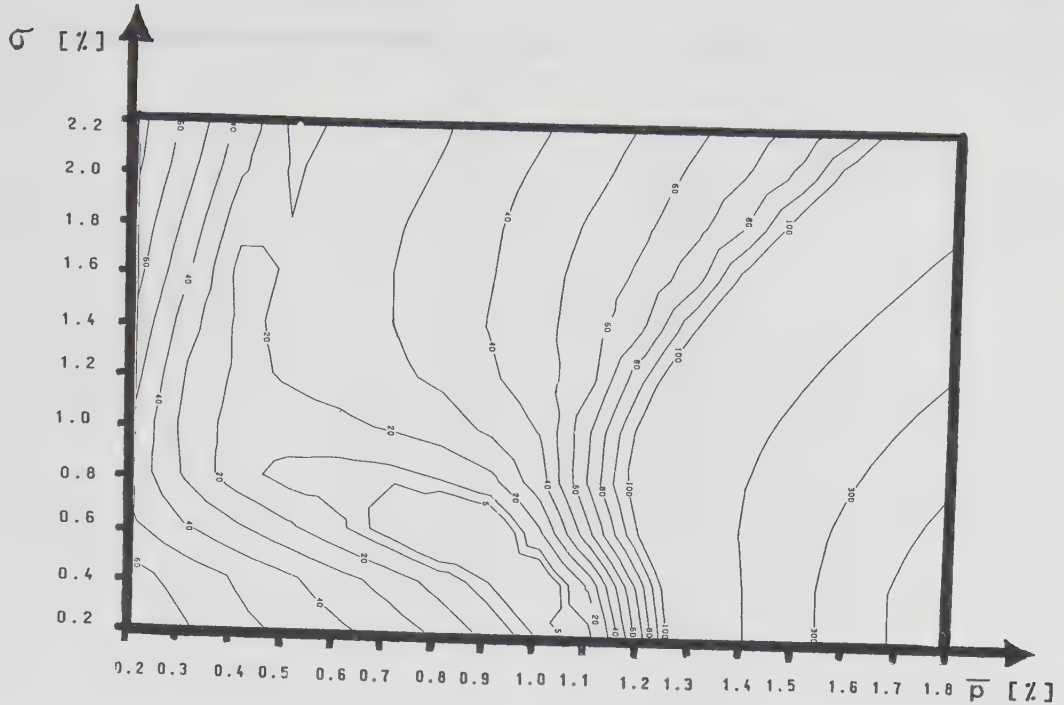

Abb. 7.5 : Isolinien von $100 * l_1(w(\bar{p}, \sigma))$
$N_L = 2000$, Trennqualitaet $p_\gamma = 1.17\%$,
Prozesskurve w : Beta-Dichte mit $\bar{p}$ und σ

Für alle Prozeßkurven w sollte daher

$$(7.10)\quad \ell_2(w) := \sum_{i=1}^{3} R((n_{A_i}, c_{A_i})|w)\, P_w(i) - \min_{i=1,2,3} R((n_{A_i}, c_{A_i})|w)$$

klein sein.

Für $h^{CU}(1) = 3.5$, $h^{CU}(2) = 3.5$ und $h^{CU}(3) = 2.2$ haben alle 3 CUSUM-Karten - einzeln geführt - im Sollwert w_S dieselbe Lauflänge: $ARL_i = 230$ $(i = 1, 2, 3)$. Abbildung 7.6 a zeigt die ARL-Isolinien für das Gesamtsystem für die oben angegebenen Werte von $h^{CU}(i)$ $(i = 1, 2, 3)$. Wir sehen, daß die Struktur der Isolinien von $\ell_1(w(\bar{p},\sigma))$ von der Struktur der ARL-Isolinien abweicht, da das den CUSUM-Karten zugrundeliegende Likelihood-Prinzip die vorgegebene Kostenstruktur ignoriert. Allerdings ist die ARL für große Werte von $\ell_1(w(\bar{p},\sigma))$ klein.

Tabelle 7.1 zeigt $P_w(i)$ $(i = 1, 2, 3)$ und $\ell_2(w(\bar{p},\sigma))$ für einige Prozeßkurven w mit $\bar{p}$ und σ. Wir sehen, daß das über $h^{CU}(1) = h^{CU}(2) = 3.5$ und $h^{CU}(3) = 2.22$ bestimmte adaptive System zu häufig die Alternative A_1 mit dem Prüfplan (50,1) wählt, wenn eigentlich die Alternative A_2 mit (125,1) kostengünstiger wäre.

Die Wahl $h^{CU}(1) = 4.0$, $h^{CU}(2) = 3.5$ und $h^{CU}(3) = 2.5$ zeigt in Abbildung 7.6 b wenig geänderte ARL-Isolinien. Jedoch zeigt Tabelle 7.1, daß diese Wahl hinsichtlich der Verlustgröße $\ell_2(w(\bar{p},\sigma))$ besser abschneidet.

Bemerkung 7.1:

Die in Abbildung 7.6 a - 7.6 b und Tabelle 7.1 angegebenen Werte für die ARL sind nicht sehr genau.

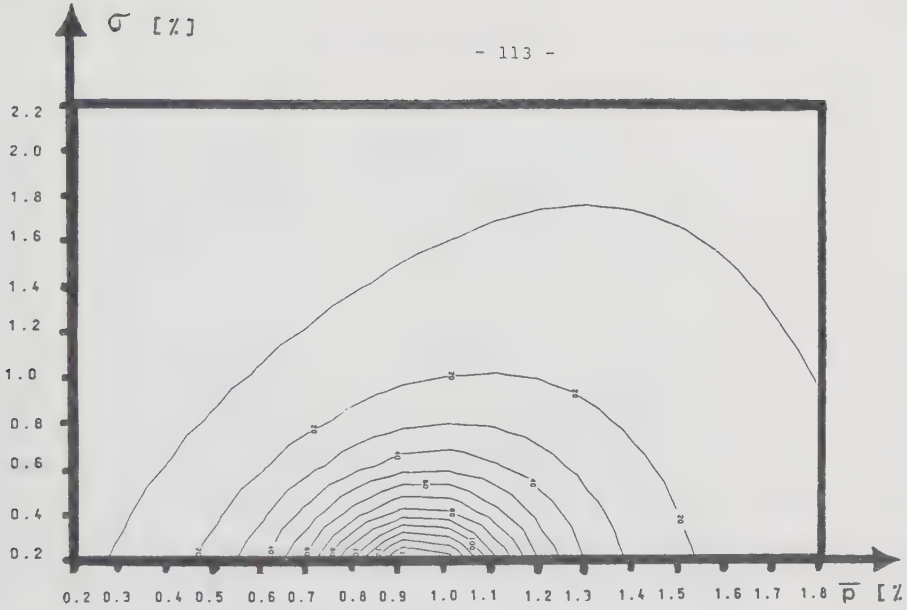

Abb. 7.6.a : ARL-Isolinien bei 3 simultan gefuehrten CUSUM-Karten zu den Alternativen A_1, A_2, A_3.
$h^{CU}(1)=3.5$, $h^{CU}(2)=3.5$, $h^{CU}(3)=2.22$

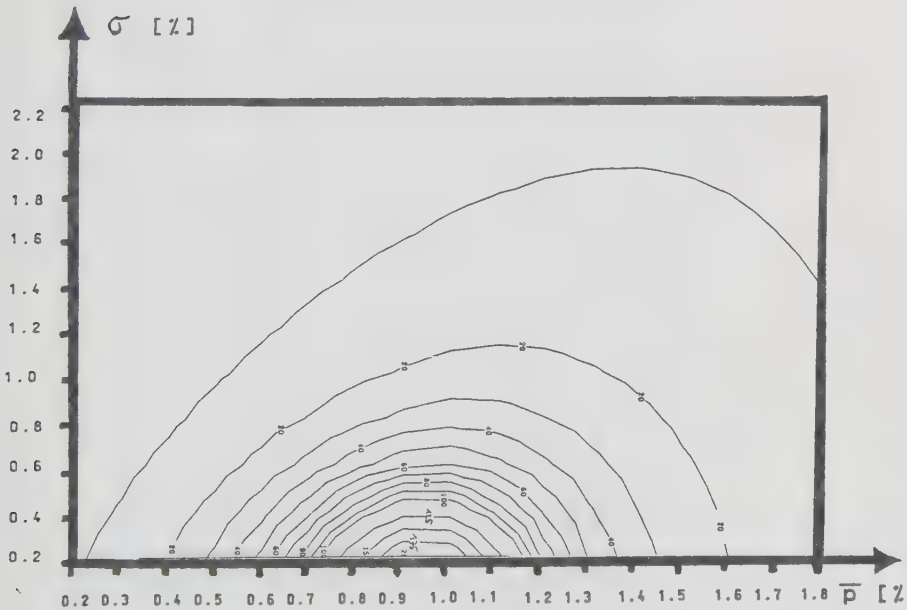

Abb. 7.6.b : ARL-Isolinien bei 3 simultan gefuehrten CUSUM-Karten zu den Alternativen A_1, A_2, A_3.
$h^{CU}(1)=4.0$, $h^{CU}(2)=3.5$, $h^{CU}(3)=2.50$

Parameter der Prozesskurve w			h(1)=h(2)=3.5 h(3)=2.22							
$\bar{p}$ in %	σ in %		Pw(1)	R([50,1]Iw)	Pw(2)	R([125,1]Iw)	Pw(3)	R([313,3]Iw)	$l_2(w)$	
0.4	0.6	(A1)	0.93	1.38	0.065	1.78	0.005	3.03	0.03	
1.0	2.2	(A2)	0.51	3.21	0.45	1.76	0.04	2.55	0.79	
1.3	1.0	(A3)	0.06	5.14	0.28	2.54	0.66	1.94	0.37	
0.6	1.2		0.73	2.42	0.24	1.79	0.03	2.91	0.47	
0.8	1.2		0.58	3.12	0.35	1.99	0.07	2.63	0.70	
1.0	1.6		0.42	3.79	0.46	2.00	0.12	2.54	0.82	
Parameter der Prozsskurve			h(1)=4.0 h(2)=3.5 h(3)=2.5							
$\bar{p}$ in %	σ in %		Pw(1)	R([50,1]Iw)	Pw(2)	R([125,1]Iw)	Pw(3)	R([313,3]Iw)	$l_2(w)$	
0.4	0.6	(A1)	0.877	1.38	0.12	1.78	0.003	3.03	0.12	
1.0	2.2	(A2)	0.35	3.21	0.61	1.76	0.04	2.55	0.53	
1.3	1.0	(A3)	0.02	5.14	0.32	2.54	0.66	1.94	0.22	
0.6	1.2		0.60	2.42	0.39	1.79	0.01	2.91	0.40	
0.8	1.2		0.43	3.12	0.52	1.99	0.07	2.63	0.53	
1.0	1.6		0.26	3.79	0.64	2.00	0.10	2.54	0.51	

Tabelle 7.1 : Werte von Pw(i) , $R([n_{A_i}, c_{A_i}]Iw)$ (i=1,2,3) und $l_2(w)$
fuer 2 verschiedene Wahlen von $h^{cv}(i)$ (i=1,2,3) .

Teilt man jedes Intervall $[0, h^{CU}(i))$ $(i = 1, 2, 3)$ in N_{Gitter} Teilintervalle, so ergeben sich N_{Gitter}^3 transiente Zustände. Bei den vorliegenden Beispielen wurde $N_{Gitter} = 15$ gewählt. Gemäß Beispiel 4.1 ist die Berechnung der ARL nur noch für niedrige Werte einigermaßen akkurat.

Die Berechnung der $P_w(i)$ $(i = 1, 2, 3)$ erfolgte für Prozeßkurven mit sehr kurzer ARL (≈ 10). In diesem Fall läßt sich die Verteilung $M^n(\vec{o}, \vec{z})$ auf den transienten Zuständen für kleine n direkt über (5.6) und (4.18) berechnen. Ist die Verteilung auf den transienten Zuständen bekannt, so läßt sich leicht die Wahrscheinlichkeit berechnen, daß eine CUSUM-Karte mit maximalem Wert die Alarmgrenze überschreitet.

□

Bemerkung 7.2:

Die Verlustgrößen $\ell_1(w(\vec{p}, \vec{c}))$ und $\ell_2(w(\vec{p}, \vec{c}))$ liefern keine eindeutige Wahl der $h^{CU}(i)$ $(i = 1, \ldots, N_A)$, wie dies bei dem ARL-Kriterium im eindimensionalen Fall geschieht.

Die Möglichkeit, die $h^{CU}(i)$ $(i = 1, \ldots, N_A)$ durch eine gemeinsame Lauflänge im Sollzustand vorzugeben, führt nicht unbedingt auf eine kostenoptimale Wahl der Parameter $h^{CU}(i)$, wie durch obige Beispiele gezeigt wurde. □

Kapitel 8

Prüfstufensteuerung des MIL-STD-105 D mit Hilfe von verallgemeinerten CUSUM-Karten

Abschnitt 8.1:

Darstellung des MIL-STD-150 D nach seiner Prüfstufensteuerung

Das Stichprobensystem MILITARY-STANDARD-105 D (MIL-STD-105 D) ist bei der attributiven Wareneingangskontrolle heute weltweit verbreitet.

Zu gegebenem Losumfang und einem vereinbarten AQL (= Acceptance Quality Level) Wert liefert die Prüfplansammlung des Systems 3 Prüfpläne (n_N, c_N), (n_R, c_R) und (n_T, c_T), die auf den 3 Prüfstufen "normale Prüfung" (Normal Inspection), "reduzierte Prüfung" (Reduced Inspection) und "verschärfte Prüfung" (Tightened Inspection) benutzt werden.

Das Verfahren startet bei normaler Prüfung und unterbricht die Warenannahme, wenn 10 Lose in verschärfter Prüfung geprüft werden (Discontinue Phase). Die Steuerung zwischen den einzelnen Prüfstufen ist in Abbildung 8.1 dargestellt.

Man erhält z. B. für einen Losumfang $N_L = 2000$ und einen AQL-Wert 0.65 %:

$(n_N, c_N) = (125,2)$

$(n_R, c_R) = (50,2)$

$(n_T, c_T) = (125,1)$

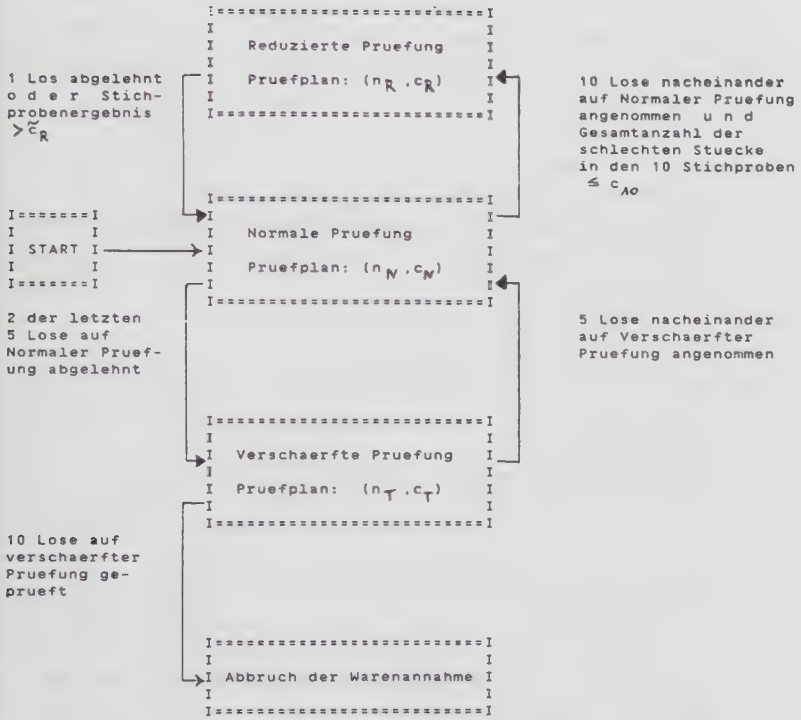

Abb. 8.1 : Die Pruefstufensteuerung des MIL-STD-105D

Der Übergang zu reduzierter Prüfung darf erfolgen, wenn die letzten 10 Lose auf normaler Prüfung angenommen wurden und bei diesen 10 Stichproben insgesamt nicht mehr $c_{10} = 4$ schlechte Stücke gefunden wurden. Der Übergang von reduzierter Prüfung zu normaler Prüfung muß erfolgen, wenn mehr als $\tilde{c}_R = 1$ schlechte Stücke in der Stichprobe sind.

Hinsichtlich der Stichprobenverteilung nehmen die Autoren des MIL-STD-105 D an, daß eine Binomial-Verteilung vorliegt und die Annahme-Wahrscheinlichkeit eines Loses bei Verwendung des Prüfplans (n, c) durch

$$(8.1) \qquad L_{n,c}(p) = \sum_{x=0}^{c} b(x|n,p) \qquad p \in (0,1)$$

berechnet werden kann, wobei p der Anteil schlechter Stücke im Los ist.

Die Prüfpläne (n_N, c_N), (n_R, c_R) und (n_T, c_T) sind so ausgewählt, daß gilt:

$$(8.2) \quad L_{n_R,c_R}(p) > L_{n_N,c_N}(p) > L_{n_T,c_T}(p) \qquad p \in (0,1)$$

Diese Prüfplanwahl unter (8.2) stellt ein System der Belohnung und Bestrafung dar: Es belohnt bei gleicher Qualität p denjenigen Produzenten, dessen Produkte mit reduzierter Prüfung geprüft werden, während der Produzent, dessen Produkte mit verschärfter Prüfung geprüft werden, durch eine verringerte Annahme-Wahrscheinlichkeit seiner Waren bestraft wird.

Der 2. Aspekt des MIL-STD-105 D liegt in seiner Ökonomie hinsichtlich des Prüfaufwandes:

Signalisieren die Stichprobenergebnisse, daß der Produktionsprozeß gut beherrscht wird und ein Qualitätsniveau $p \leq AQL$ erreicht wird, so wird der Prüfaufwand wegen $n_R < n_N$ erheblich reduziert.

Signalisieren die Stichprobenergebnisse ein mögliches Überschreiten des AQL-Werts, so wird der Stichprobenaufwand wegen $n_R < n_N \leq n_T$ erhöht. Die Stichprobe wird damit informativer hinsichtlich des Verhaltens von p.

Der Nachteil bei der Anwendung des MIL-STD-105 D besteht in seiner komplizierten Prüfstufensteuerung. So müssen stets die letzten 10 Stichprobenergebnisse vorliegen, die nach jedem Prüfvorgang aufdatiert werden müssen.

Eine gewisse Reduktion des Aufwands erreicht man, indem man die Regel über die maximale Gesamtanzahl der schlechten Stücke beim Übergang von normaler zu reduzierter Prüfung ignoriert. In diesem Fall kann man seinen Informationsaufwand auf die 10 letzten Annahme-/Ablehnungs-Entscheidungen reduzieren.

Mit hohen Risiken ist jedoch eine Praxis verbunden, die Wareneingangsprüfung nur mit (n_N, c_N) durchzuführen. Zum Beispiel beträgt die Annahme-Wahrscheinlichkeit beim Plan (125,2) bei $p = 2.14\,\% \approx 3$ AQL immer noch 0.5!

Abschnitt 8.2:

Prüfstufensteuerung mit Hilfe von verallgemeinerten CUSUM-Karten

Das verallgemeinerte CUSUM-Schema bietet die Möglichkeit, die Prüfstufensteuerung des MIL-STD-105 D auf das Führen einer einzigen (verallgemeinerten) CUSUM-Karte zu reduzieren.

Der Sollzustand der Produktion sei durch p_S = AQL gegeben. Als Alarmzustand p_A, der möglichst schnell aufgedeckt werden soll und zu einem Abbruch der Warenannahme führt, definieren wir die Qualität, die unter normaler Prüfung gerade noch mit Wahrscheinlichkeit 0.5 angenommen wird.

Wir sind damit in der Lage, ein verallgemeinertes CUSUM-Schema GCU_t der Form (2.17) mit 3 Prüfstufen anzuwenden.

Hierbei findet die reduzierte Prüfung zum Zeitpunkt t Anwendung, falls $GCU_{t-1} < h_1^{CU} = 0$ ist. Die normale Inspektion (d. h. Prüfen mit Prüfplan (n_N, c_N) findet zum Zeitpunkt t statt, falls $0 \leq GCU_{t-1} < h_2^{CU} =: h_T$. Die verschärfte Prüfung zum Zeitpunkt t wird bei $h_T \leq GCU_{t-1} < h^{CU} =: h_D$ angewendet. Überschreitet GCU_t den Wert $h^{CU} = h_D$, so tritt die Discontinue-Regelung in Kraft.

Die Bestimmung der jeweiligen Prüfstufe zum Zeitpunkt t erfolgt bei diesem Verfahren also durch den vorhergehenden Wert von GCU_{t-1}. Der Speicheraufwand hinsichtlich der vergangenen Stichprobenergebnisse reduziert sich damit auf das Merken einer einzigen Zahl!

Zur Festlegung der Prüfstufensteuerung wird lediglich die Angabe der beiden Werte h_T und h_D benötigt.

Da die Autoren des MIL-STD-105 D für die Verteilung der X_t $(t \geq 1)$ die Binomialverteilung unterstellt haben, läßt sich in Anwendung von Beispiel 2.1 das GCU_t-Schema transformieren auf:

$$GK_0 = 0$$

$$(8.3)\quad GK_t = \max\{0, GK_{t-1}\} + \begin{cases} X_t - k(n_R) & GK_{t-1} < 0 \\ X_t - k(n_N) & 0 \leq GK_{t-1} < h_T^K \\ X_t - k(n_T) & h_T^K \leq GK_{t-1} < h_D^K \end{cases}$$

$$t \geq 1$$

und Abbruch der Warenannahme, falls:

$$(8.4)\qquad GK_t \geq h_D^K \qquad t \geq 1$$

Für $N_L = 2000$ und $AQL = p_S = 0.65\ \%$ erhalten wir $n_R = 50$, $n_N = n_T = 125$ und $p_A = 2.14\ \%$. (2.20) liefert die Werte:

$$k(n_R) = k(50) = 0.63$$

$$k(n_N) = k(n_T) = k(125) = 1.56$$

Mit Hilfe des CUSUM-Prinzips lassen sich weitere CUSUM-Versionen des MIL-STD-105 D formulieren. Dem ursprünglichen MIL-STD-105 D am nächsten ist ein CUSUM-Schema, das nur auf der Annahme-/Ablehnungs-Entscheidung der Lose basiert. In diesem Fall sind die beobachteten Zufallsgrößen $X2_t$ (0-1)-Variablen. Wir definieren $X2_t$ auf den einzelnen Prüfstufen R, N und T wie folgt:

Auf Prüfstufe s = R:

$$X2_t := \begin{cases} 1 & X_t \leq \tilde{c}_R \quad (\text{"Verbleiben in red. Prüfung"}) \\ 0 & X_t > \tilde{c}_R \end{cases}$$

Auf Prüfstufe s = N:

$$X2_t := \begin{cases} 1 & X_t \leq c_N \quad \text{("t. Los auf normaler Prüfung angenommen")} \\ 0 & X_t > c_N \end{cases}$$

Auf Prüfstufe s = T:

$$X2_t := \begin{cases} 1 & X_t \leq c_T \quad \text{("t. Los auf verschärfter Prüfung abgelehnt")} \\ 0 & X_t > c_T \end{cases}$$

Die Zuwachsfunktionen $Y2_t(s)$ s = R, N, T des verallgemeinerten CUSUM-Schemas (2.17) haben damit die folgende Gestalt:

$$(8.5) \quad Y2_t(\sigma) = \ln \frac{f(X2_t|\sigma, p_A)}{f(X2_t|\sigma, p_S)} \qquad \sigma = R, N, T$$

Für N_L = 2000 und AQL = p_S = 0.65 % und p_A = 2.14 % erhält man für n_R = 50, $n_N = n_T$ = 125, $\tilde{c}_R$ = 1, c_N = 2 und c_T = 1 unter Benutzung von (8.1):

$$(8.6) \quad \begin{aligned} &P(X2_t=1|s = R, p = p_S) = 0.95 \\ &P(X2_{t=1}|s = R, p = p_A) = 0.70 \end{aligned}$$

$$(8.7) \quad \begin{aligned} &P(X2_t=1|s = N, p = p_S) = 0.95 \\ &P(X2_t=1|s = N, p = p_A) = 0.50 \end{aligned}$$

$$(8.8) \quad \begin{aligned} &P(X2_t=1|s = T, p = p_S) = 0.70 \\ &P(X2_t)=1|s = T, p = p_A) = 0.25 \end{aligned}$$

Wir erhalten damit bei der verallgemeinerten CUSUM-Karte (2.17) die Zuwachsfunktionen $Y2_t(s)$ auf den Prüfstufen s = R, N, T.

Wert der Zuwachsfunktion $Y2_t(s)$ bei :		
Pruefstufe	$X2_t = 1$	$X2_t = 0$
s=R	-0.3	1.8
s=N	-0.6	2.3
s=T	-1.0	0.9

Schließlich können wir die Varianz der Prozeßkurve bei der Prüfstufensteuerung mitberücksichtigen.

Die Autoren des MIL-STD-105 D haben als Stichprobenverteilung die Binomial-Verteilung zugrundegelegt. Wie wir über Satz 6.2 schließen können, impliziert dies die Annahme einer Prozeßkurve mit Varianz $\sigma^2 = 0$, was wenig realistisch ist.

Bemerkung 8.1:

Die dem MIL-STD-105 D zugrunde liegende Stichprobenverteilung wird häufig als bedingte Verteilung von $X^S = X^S(n, N_L)$ gegeben $X^L = X^L(N_L) = pN_L$ interpretiert, wobei die Hypergeometrische Verteilung durch eine Binomial-Verteilung approximiert wird. In diesem Fall können keine Rückschlüsse auf den Erwartungswert und die Varianz der unbedingten Verteilung von X^S gezogen werden. Allenfalls läßt sich noch schlüssig argumentieren, daß der Erwartungswert $\bar{p}$ im Sollzustand kleiner als AQL sein muß. □

Wir wollen annehmen, daß der Sollzustand durch eine Prozeßkurve mit:

$\bar{p}_S$ = AQL = 0.65 %

σ_S = AQL = 0.65 %

beschrieben sei. Als mögliche Alarmzustände, die zu einem Abbruch der Warenannahme führen sollen, betrachten wir:

(A MIL1) $\bar{p}_A$ = 2.14 % $\quad \sigma_A = \sigma_S$ = 0.65 %

(A MIL2) $\bar{p}_A$ = 2.14 % $\quad \sigma_A$ = 2.14 %

Auf den Prüfstufen s = R, N, T hat dann X_t unter einer Prozeßkurve mit $\bar{p}$ und σ eine Polyaverteilung mit:

$$(8.9) \qquad f(x|s, \bar{p}, \sigma) = \begin{cases} f(x|n_R, \bar{p}, \sigma) & s = R \\ f(x|n_N, \bar{p}, \sigma) & s = N \\ f(x|n_T, \bar{p}, \sigma) & s = T \end{cases}$$

Hierbei ist $f(X|n, \bar{p}, \sigma)$ durch (6.5) gegeben. Die Zuwachsfunktionen $Y3_t(s)$ s = R, N, T des verallgemeinerten CUSUM-Schemas (2.17) sind in diesem Fall gegeben durch:

$$(8.10) \qquad Y3_t(\sigma) = \begin{cases} \ln \dfrac{f(X_t|n_R, \bar{p}_A, \sigma_A)}{f(X_t|n_R, \bar{p}_S, \sigma_S)} & \sigma = R \\ \ln \dfrac{f(X_t|n_N, \bar{p}_A, \sigma_A)}{f(X_t|n_N, \bar{p}_S, \sigma_S)} & \sigma = N \\ \ln \dfrac{f(X_t|n_T, \bar{p}_A, \sigma_A)}{f(X_t|n_T, \bar{p}_S, \sigma_S)} & \sigma = T \end{cases}$$

Abschnitt 8.3:

Die Berechnung der Kenngrößen von MIL-STD 105 D und seiner CUSUM-Versionen

Wir wollen nun den MIL-STD-105 D und dessen CUSUM-Versionen daraufhin untersuchen, wie gut sie Soll- und Alarmzustand erkennen, d. h. wir wollen die Lauflänge N bis zum Erreichen der "Discontinue Phase" berechnen.

Untersuchungen für den MIL-STD-105 D existieren bisher nur für den wenig realistischen Fall, daß $p_t = p$ $(t \geq 1)$ gilt (vgl. Brown and Rutemiller (1973) , Koyama (1974), Koyama (1978) und Hald (1981), p. 288 - 298).

Um die Ergebnisse der Abschnitte 4.1 - 4.5 benutzen zu können, müssen wir die Übergangsdichte auf den transienten Prüfzuständen des jeweiligen Prüfverfahrens bestimmen.

Bemerkung 8.2:

Wir wollen den MIL-STD-105 D in einer modifizierten Form anwenden. Die Bedingungen für den Übergang von normaler nach reduzierter Prüfung, wonach die Gesamtanzahl der schlechten Stücke in den 10 letzten Stichproben nicht mehr als c_{10} betragen darf, wird in der Praxis häufig ignoriert. Da diese Bestimmung auch mathematisch sehr unhandlich ist, wird im folgenden diese Regel ignoriert, so daß allein 10 nacheinander auf normaler Prüfung angenommene Lose ausreichen, um in die reduzierte Prüfung zu gelangen.

Eine Darstellung, wie die Übergangswahrscheinlichkeiten zwischen den Prüfzuständen des nicht-modifizierten MIL-STD-105 D bestimmt werden kann, findet man bei Brown and Rutemiller (1973). □

Abbildung 8.2a zeigt die 24 transienten Prüfzustände des MIL-STD-105 D und die zugehörigen Übergangswahrscheinlichkeiten (vgl. Koyama (1974)). Die Übergangsdichte M_{MIL} zwischen den transienten Prüfzuständen hängt dabei nur von 3 Werten P_R, P_N und P_T ab.

Satz 8.1:

Die Übergangswahrscheinlichkeiten zwischen den transienten Prüfzuständen des (modifizierten) MIL-STD-105 D sind durch die in Abbildung 8.2a angegebenen Werte bestimmt. Unter einer Prozeßkurve w(p) mit Standardabweichung σ sind P_R, P_N und P_T gegeben durch:

$$(8.11)\quad P_R := \begin{cases} \sum_{x=0}^{c_R} \int_0^1 b(x|n_R, p)\, w(p)\, dp & \sigma > 0 \\ \sum_{x=0}^{c_R} b(x|n_R, p) & \sigma = 0 \end{cases}$$

$$(8.12)\quad P_N := \begin{cases} \sum_{x=0}^{c_N} \int_0^1 b(x|n_N, p)\, w(p)\, dp & \sigma > 0 \\ \sum_{x=0}^{c_N} b(x|n_N, p) & \sigma = 0 \end{cases}$$

$$(8.13)\quad P_T := \begin{cases} \sum_{x=0}^{c_T} \int_0^1 b(x|n_T, p)\, w(p)\, dp & \sigma > 0 \\ \sum_{x=0}^{c_T} b(x|n_T, p) & \sigma = 0 \end{cases}$$

Beweis:

Es gilt analog zum Beweis von Satz 6.1 für eine Prozeßkurve w(p) mit $\sigma > 0$:

$$\begin{aligned} P_N &= P(\text{"Annahme eines Loses auf normaler Prüfung"}) \\ &= \sum_{X=0}^{N_L} \Big(\sum_{x=0}^{c_N} h(x|n_N, X, N_L) \Big) P\big(X^L(N_L) = X\big) \\ &= \sum_{x=0}^{c_N} \int_0^1 b(x|n_N, p)\, w(p)\, dp \end{aligned}$$

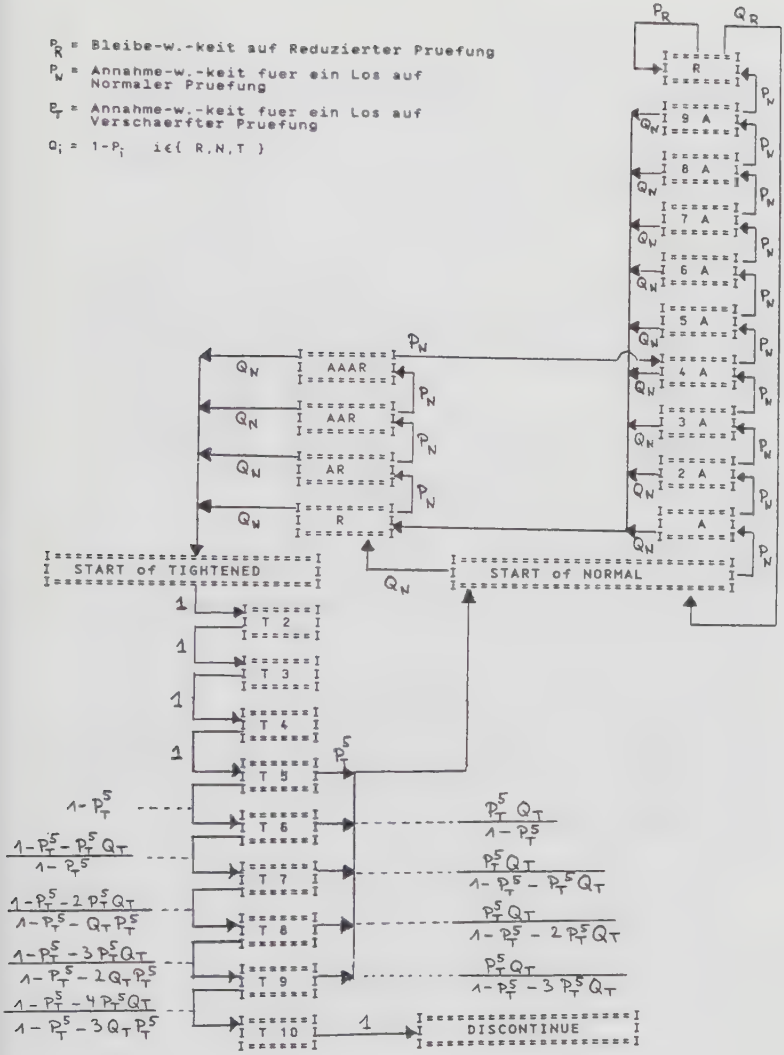

Abb. 8.2.a : Die Pruefzustaende des MIL-STD-105D und die Uebergangswahrscheinlichkeiten zwischen ihnen .

Damit erhält man z. B.:

$$P(\text{"A"} \mid \text{"Start of Normal Insp."}) = P_N$$

Die anderen Übergangswahrscheinlichkeiten auf normaler und reduzierter Prüfung berechnen sich analog.

Während der verschärften Prüfung bezeichne $T_i XA$ ($i = 2, 3, \ldots, 9$) ($X = 1, 2, 3, 4$) einen Prüfzustand, der dadurch charakterisiert ist, daß vor der i-ten Prüfung auf verschärfter Prüfung die letzten X Lose angenommen wurden.

Der Zustand $T_i R$ ($i = 2, 9$) bezeichne den Prüfzustand, daß vor der i-ten Prüfung das letzte Los abgelehnt wurde.

Die Übergangswahrscheinlichkeiten zwischen diesen Prüfstufen sind durch P_T bestimmt und in Abbildung 8.2.b wiedergegeben.

Wir fassen nun die Prüfzustände $T_i XA$ ($X = 1, 2, 3, 4$) und $T_i R$ für jedes i ($i = 2, \ldots, 9$) zu einem Prüfzustand T_i zusammen. Analog zu Koyama (1974) erhält man für die Übergänge von T_i ($i = 2, \ldots, 9$) zum Prüfzustand "Start of Normal Insp." :

$$\begin{aligned} P(\text{"Start of Normal Insp."} \mid T_5) &= P(\text{"Alle 5 Lose auf T. Insp. angenommen"}) \\ &= P_T^5 \end{aligned}$$

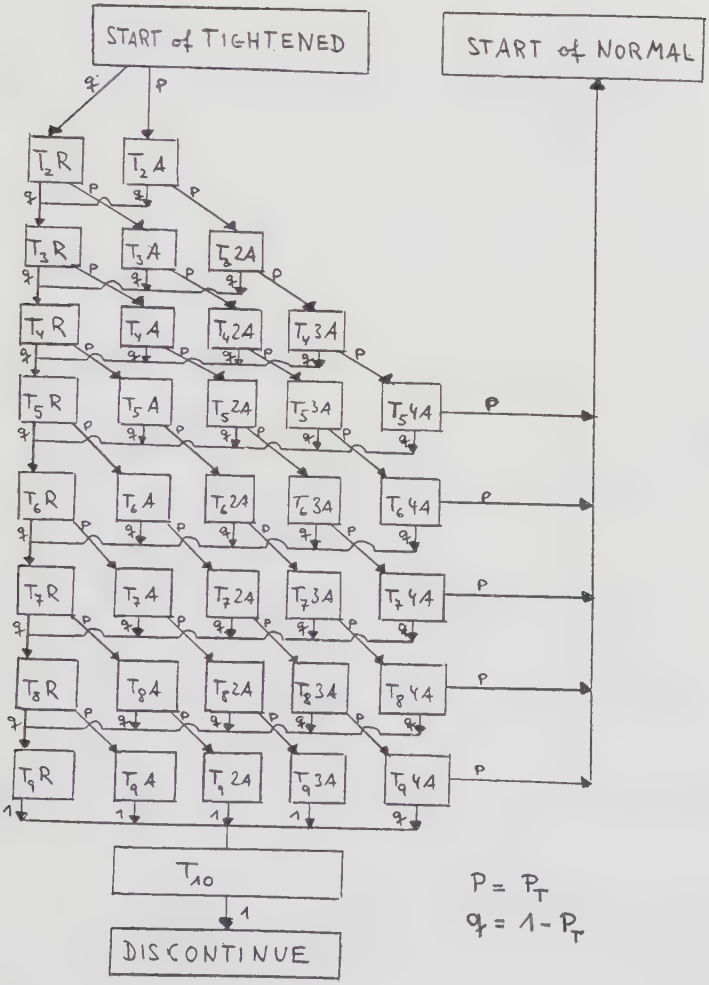

Abb. 8.2.b : Die Pruefzustaende des MIL-STD-105D bei Verschaerfter Pruefung und die Uebergangswahrscheinlichkeiten zwischen ihnen .

$$P(\text{"Start of Normal Insp."} \mid T_6)$$

$$= \frac{P(\text{"1. Los auf T.Insp. abgelehnt, danach 5 Lose ang."})}{P(\text{"Bei } T_5 \text{ nicht nach "Start of N. Insp."})}$$

$$= \frac{(1 - P_T)P_T^5}{1 - P_T^5}$$

Die anderen Übergangswahrscheinlichkeiten berechnet man analog. □

Satz 8.2:

M_{MIL} erfüllt die Voraussetzung (*) von Theorem 4.1.

Beweis:

Von jedem transienten Prüfzustand E_i $(i = 1, \ldots, 24)$ kann der Zustand E_1 = "Prüfen mit reduzierter Prüfung" in $m_{i,1} \leq 15$ Schritten erreicht werden.

Von E_1 kann jeder andere Prüfzustand E_j in $m_{1,j} \leq 11$ Schritten erreicht werden.

Die Wahrscheinlichkeit, daß das System in E_1 verweilt, ist positiv.

Damit läßt sich analog zum Beweis von Satz 4.2 zeigen, daß M_{MIL}^{26} positiv ist. □

Die Übergangsdichte M_{CUMIL} der CUSUM-Version des MIL-STD-105 D erfüllt die Voraussetzung (*) von Theorem 4.1, falls die Voraussetzungen (4.25) und (4.26) von Satz 4.2 erfüllt sind.

Voraussetzung (4.26) ist für alle im Abschnitt 8.2 genannten CUSUM-Versionen des MIL-STD-105 D erfüllt.

Die Bedingung (4.25), nach der alle transienten Prüfzustände vom Zustand E_{-1} in endlich vielen Schritten erreicht werden können, überprüft man durch Berechnung von $M^n_{CUMIL}(-1, z)$, $z \in \text{supp } \mu$ $n = 1, 2, \ldots$. Bei Benutzung des in Abschnitt 4.1 geschilderten Rundungsverfahrens wird (4.25) für alle CUSUM-Versionen des MIL-STD-105 D erfüllt.

Abschnitt 8.4:

Anpassen der Parameter der CUSUM-Versionen des MIL-STD-105 D

Um den MIL-STD-105 D und seine CUSUM-Versionen miteinander vergleichen zu können, müssen noch die Parameter h_T und h_D gewählt werden.

Als Stichprobenverteilung legen wir dabei die von den Autoren des MIL-STD-105 D benutzte Binomial-Verteilung zugrunde.

Die Wahl der Parameter h_T und h_D erfolgt unter dem Ziel, die Verteilung der Lauflänge N unter p = AQL für alle Verfahren möglichst gut anzupassen.

Wir wählen zunächst h_T so, daß die mittlere Laufzeit bis zum Erreichen der verschärften Prüfung $ARL_{N\rightarrow T}$ für alle Verfahren übereinstimmt. Dabei soll die reduzierte Prüfung nicht mit berücksichtigt werden. Für den MIL-STD-105 D ist $ARL_{N\rightarrow T}$ gegeben durch die mittlere Dauer, bis 2 von 5 Losen unter normaler Prüfung abgelehnt werden. Es gilt (vgl. Hald (1981), S. 88):

$$(8.14) \quad ARL^{MIL}_{N\rightarrow T} = \frac{2 - P_N^4}{(1-P_N)(1-P_N^4)}$$

$ARL_{N\rightarrow T}$ ist bei den CUSUM-Versionen isoton von h_T abhängig. Da die Zuwachsgröße $Y_t(s = N)$ diskret ist, läßt sich die Forderung

$$(8.15) \quad ARL^{MIL}_{N\rightarrow T} = ARL^{CUMIL}_{N\rightarrow T}$$

nicht immer exakt erfüllen.

Die Wahl von h_D erfolgt dann so, daß die mittlere Laufzeit bis zum Erreichen der Discontinue Phase ARL = ARL_D unter Berücksichtigung aller Prüfstufen möglichst gut übereinstimmt.

Für N_L = 2000 und p = AQL = 0.65 % erhält man:

(8.16) $\quad ARL^{MIL}_{N \rightarrow T} = 134 \qquad ARL^{MIL}_{D} = 953$

Tabelle 8.1 zeigt die dazugehörigen Werte von h_T und h_D für die in Abschnitt 8.2 erwähnten 4 alternativen CUSUM-Versionen des MIL-STD-105 D.

CUSUM-Version	h_T	$ARL_{N \rightarrow T}$	h_D	ARL_D
Binomial Version				
p_S = 0.65%	3.1	138	4.73	973
p_A = 2.14%				
Bernoulli Version				
vgl (8.5)	2.8	134	4.73	931
Polya Version				
$\bar{p}_S$ = 0.65% σ_S = 0.65%	2.8	137	4.3	920
$\bar{p}_A$ = 2.14% σ_A = 0.65%				
Polya Version			2.55	920
$\bar{p}_S$ = 0.65% σ_S = 0.65%	1.45	135	2.56	1050
$\bar{p}_A$ = 2.14% σ_A = 2.14%				

Tabelle 8.1 : Anpassung der Werte von h_T und h_D von 4 verschiedenen CUSUM-Versionen des MIL-STD-105D zu den vorgegeben Werten $ARL_{N \rightarrow T}$=134 u. ARL_D =953 beim MIL-STD-105D unter p=AQL=0.65%

Abschnitt 8.5:

Vergleich der Lauflängenverteilung von MIL-STD-105 D und seiner CUSUM-Versionen

Wir wollen zunächst die Verteilung der Lauflänge N bis zum Erreichen der Discontinue Phase unter der Annahme bestimmen, daß die Stichprobenergebnisse identisch binomial-verteilt sind. Dies entspricht dem bisher in der Literatur behandelten Fall.

Zunächst vergleichen wir für p = 0.65 % = AQL, 0.8 % = 1.25 AQL, 1.3 % = 2 AQL und 2 % ≈ 3 AQL die wichtigsten Kenngrößen der Lauflängenverteilung. Neben den Erwartungswerten $ARL_{N \to T}$ und ARL_D gibt Tabelle 8.2 die Standardabweichung σ_D der Lauflänge N, ihren 50 %-Punkt $N_{.50}$ und ihren 90 %-Punkt an. Nach Satz 4.3 gilt $P(N = n + 1) / P(N = n) \to \lambda$.

Die genannten Werte sind in Tabelle 8.2 für den MIL-STD-105 D und die Binomial-CUSUM-Version angegeben.

MIL-STD-105D

p in %	λ	$ARL_{N \to T}$	ARL_D	σ_D	N.50	N.90
0.65	0.99896	134	953	959	658	2204
0.80	0.9954	57.1	220	219	151	507
1.30	0.940	11.5	25.6	14.5	19.5	44.2
2.00	0.775	4.5	14.5	2.8	13.5	18.5

Binomial-CUSUM-Version des MIL-STD-105D

p in %	λ	$ARL_{N \to T}$	ARL_D	σ_D	N.50	N.90
0.65	0.99897	138	973	971	674	2237
0.80	0.9959	54.4	248	246	172	570
1.30	0.940	9.6	19.4	15.5	14.5	40.0
2.00	0.735	3.6	5.5	3.4	4.5	9.5

Tabelle 8.2 : Vergleich der Kenngroessen von MIL-STD-105D und der Binomial-CUSUM-Version des MIL-STD-105D unter einer Binomial-Verteilung mit Parameter p .

Wir sehen, daß die Anpassung im Erwartungswert bei p = AQL auch eine gute Anpassung in den übrigen Parametern der Verteilung liefert. Für p = 1.3 % ist ARL_D bei der CUSUM-Version um 25 % kürzer als beim MIL-STD-105 D und bei p = 2 % um 62 %.

Die Abbildungen 8.3 a - 8.3 d vergleichen die gesamte Verteilung von N für p = AQL, 1.25 AQL, 2 AQL und 3 AQL. Trotz der sehr unterschiedlichen Formen der Dichte von N für $N \leq 30$ bei p = AQL und p = 1.25 AQL hat dies kaum Auswirkungen auf die übrigen Kenngrößen der Verteilung von N. Bei p = 2 % zeigt sich eine deutliche Überlegenheit der CUSUM-Version, einen schlechten Qualitätszustand zu erkennen.

Auffällig ist, daß alle Dichten schon für relativ kleine n ($n \geq 30$) die asymptotische Beziehung $P(N=n+1) = \lambda\, P(N=n)$ erfüllen, d. h. daß alle Dichten für $n \geq 30$ gut durch eine geometrische Verteilung approximiert werden können. Für große Lauflängen wird man daher erwarten können, daß die Approximation der Verteilung der Lauflänge durch

$$P(N = n) \approx P_{Geo}(n) := (1 - \lambda)\,\lambda^n \qquad n = 0, 1, 2, \ldots \tag{8.17}$$

zumindest hinsichtlich der Kenngrößen ARL_D, σ_D und $N_{0.90}$ eine gute Approximation liefert.

Die Kenngrößen der entsprechenden geometrischen Verteilung sind gegeben durch:

$$E_{Geo}(N) = \frac{\lambda}{1-\lambda}\,, \quad \sigma_{Geo} = \frac{\sqrt{\lambda}}{1-\lambda} \tag{8.18}$$

$$N_{\alpha}^{Geo} = \frac{\ln(1-\alpha)}{\ln \lambda} - 1$$

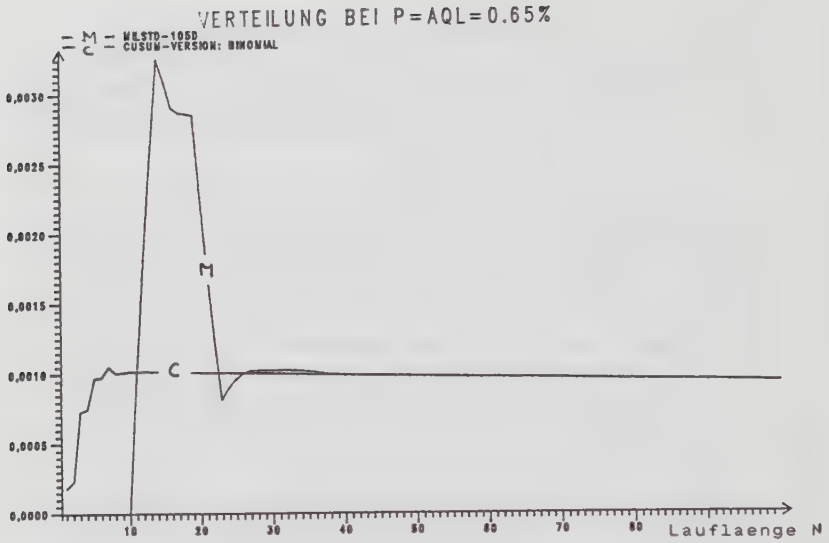

Abb. 8.3.a : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Binomial mit p=0.65%

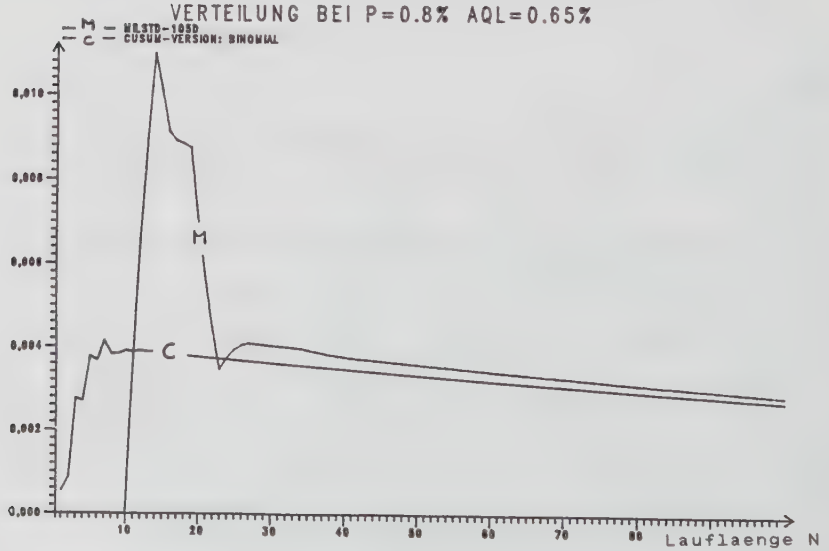

Abb. 8.3.b : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Binomial mit p=0.8%

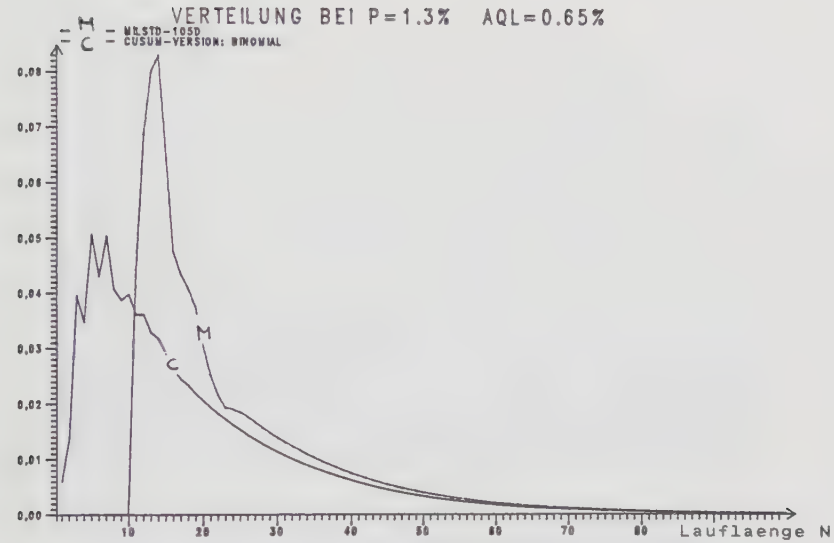

Abb. 8.3.c : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : $h_{\neg}$ =3.1 $h_{\supset}$ =4.73
Stichprobenverteilung : Binomial mit p=1.3%

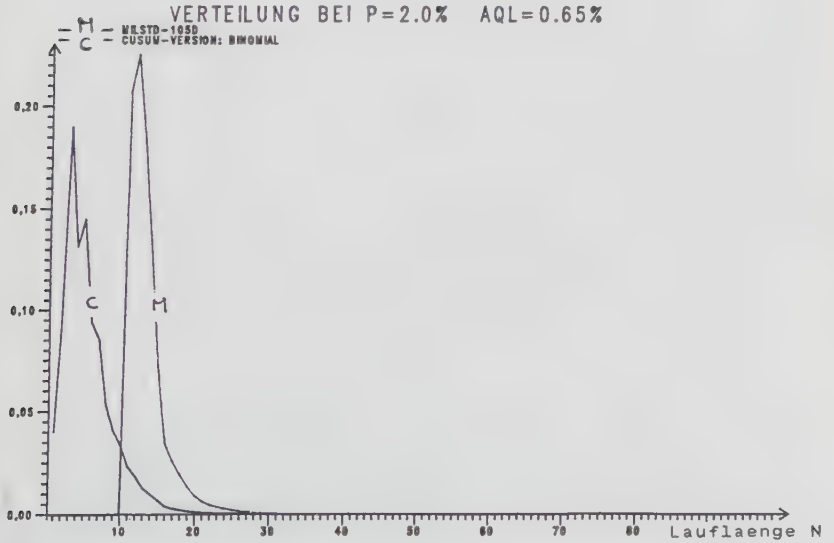

Abb. 8.3.d : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : $h_{\neg}$ =3.1 $h_{\supset}$ =4.73
Stichprobenverteilung : Binomial mit p=2.0%

Durch die Wahl:

$$\hat{\lambda} = \frac{ARL_D}{ARL_D + 1} \tag{8.19}$$

wird $ARL_D = E_{Geo}(N)$ gesichert.

In Tabelle 8.3 wird die Approximation von ARL_D, σ_D, $N_{.90}$ und $\lambda_\infty := \lim_{n\to\infty} \frac{P(N = n+1)}{P(N = n)}$ beim MIL-STD-105 D durch die entsprechenden Werte einer geometrischen Verteilung mit $\hat{\lambda}$ für verschiedene Werte von p dargestellt. Wir sehen, daß für $ARL_D \geq 50$ sowohl die Standardabweichung als auch der 90 %-Punkt gut durch $\sigma_{Geo}(\hat{\lambda})$ und $N^{Geo}_{.90}(\hat{\lambda})$ approximiert werden.

p in %	ARL_D	λ_∞	$\hat{\lambda}$	σ_D	σ_{Geo}	$N^D_{.90}$	$N^{Geo}_{.90}$
0.65	953	0.99896	0.99895	959	960	2204	2190
0.70	550	0.99820	0.99818	554	549	1272	1263
0.80	220	0.9954	0.9954	219	217	507	498
0.90	108	0.9905	0.9908	104	108	243	248
1.00	63	0.9828	0.9843	57	63	137	144
1.10	42	0.9721	0.9765	34	42	84	95
1.20	31	0.958	0.968	22	31	60	69
1.30	26	0.94	0.96	15	28	44	55

Tabelle 8.3 : Approximation der Lauflaengenverteilung des MIL-STD-105D durch eine geometrische Verteilung mit gleichem Erwartungswert . Berechnung der Lauflaenge unter einer Binomial Verteilung mit Parameter p .

Bemerkung 8.3:

Das obige Ergebnis reduziert den Rechenaufwand hinsichtlich der Kenngrößen der Lauflängenverteilung auf die Berechnung von ARL, falls $ARL \geq 50$. Hierfür ist in Theorem 4.11 ein effektiver Algorithmus angegeben.

Analoge Resultate gelten auch für die CUSUM-Versionen des MIL-STD-105 D.

Allerdings zeigt (8.18) auch eine Schwäche dieser Verfahren auf. Da $\hat{\lambda}$ nahe bei 1 ist, gilt:

$$(8.20) \qquad \sigma_D \approx \sigma_{Geo}(\hat{\lambda}) = \frac{\hat{\lambda}^{1/2}}{1-\hat{\lambda}} \approx \frac{\hat{\lambda}}{1-\hat{\lambda}} = ARL_D$$

D. h. eine Vergrößerung der ARL im Sollzustand zieht immer eine Vergrößerung der Streuung von N nach sich. □

Wir wollen jetzt den realistischeren Fall diskutieren, daß die Prozeßvarianz σ^2 größer als Null ist.

Der Einfluß der Prozeßvarianz σ^2 bei $\bar{p} = AQL = 0.65\ \%$ auf die Kenngrößen der Lauflängenverteilung von MIL-STD-105 D und der Binomial-CUSUM-Version wird in Tabelle 8.4 wiedergegeben. Bemerkenswert ist, daß der MIL-STD-105 D auftretende Prozeßvarianzen zunächst durch eine kürzere ARL bestraft. Für $\sigma > 1.0\ \%$ werden noch größere Prozeßvarianzen jedoch wieder durch eine Verlängerung der ARL belohnt. Die Binomial-CUSUM-Version bestraft im angegebenen Bereich gleichmäßig höhere Prozeßvarianzen.

Die Streuung σ_D der Lauflänge ist in allen Fällen ungefähr gleich dem Erwartungswert ARL_D. Wir können daraus schließen, daß auch bei Vorliegen einer Prozeßvarianz die Verteilung der Lauflänge gut durch eine geometrische Verteilung approximiert wird.

	MIL-STD-105D					Binomial-CUSUM-Version des MIL-STD-105D				
σ in %	λ	ARL_D	σ_D	N.50	N.90	λ	ARL_D	σ_D	N.50	N.90
0.0	0.99896	953	956	658	2204	0.99897	973	971	674	2237
0.3	0.99805	509	513	351	1179	0.99763	422	421	292	970
0.7	0.9954	219	217	151	504	0.9920	125	125	86.1	288
1.0	0.9944	183	180	127	419	0.9871	76.5	70.0	52.0	176
1.2	0.9946	189	186	131	433	0.9838	61.0	61.0	41.0	140
1.4	0.9951	210	206	145	479	0.9816	53.0	53.5	35.6	122
1.6	0.9957	237	234	165	543	0.980	48.5	49.3	32.5	112
2.2	0.9978	456	454	316	1049	0.976	41.5	41.0	27.2	84.0

Tabelle 8.4 : Einfluss der Prozessvarianz σ^2 auf die Kenngroessen von MIL-STD-105D und der Binomial-CUSUM-Version des MIL-STD-105D. Stichprobenverteilung : Polya mit $\bar{p}$ =AQL=0.65% und σ .

Die Abbildungen 8.4 a - 8.4 d zeigen den Verlauf der Dichte der Lauflänge N bei $\bar{p}$ = AQL = 0.65 % und σ = 0.3 %, 0.7 %, 1.0 % und 2.2 %. Verglichen wird die Dichte der Lauflänge des MIL-STD-105 D und der Binomial-CUSUM-Version. Für $\sigma \geq 0.7$ % sehen wir den Einfluß der "Ausreißer"-Stichproben, die den Prozeß der Warenannahme bei der CUSUM-Version schon nach einem Los abbrechen lassen. Dies ist unter normaler und verschärfter Prüfung bei n = 125 und h_D = 4.73 für $X \geq 6$ der Fall.

Die Abbildungen 8.5 a und 8.5 b zeigen die ARL_D-Isolinien vom MIL-STD-105 D und der Binomial-CUSUM-Version auf einem $(\bar{p}, \sigma)$-Gitter.

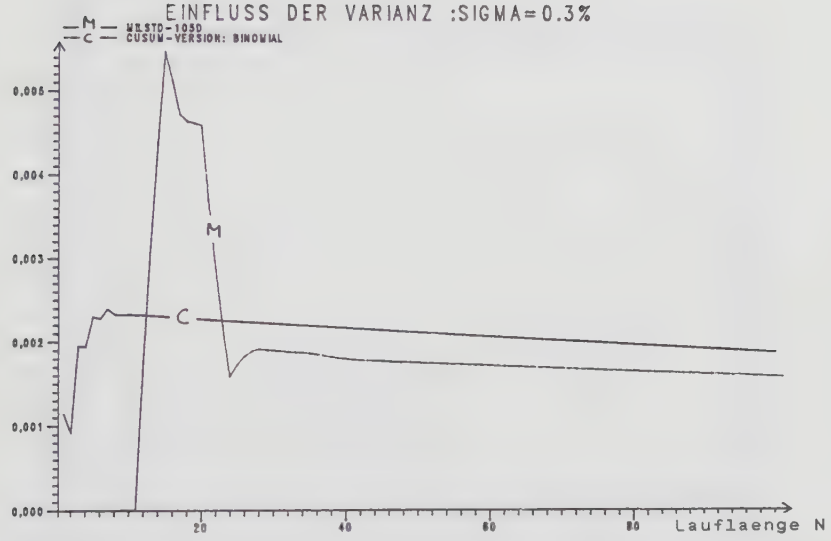

Abb. 8.4.a : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase unter einer Prozesskurve mit $\bar{p}$ =AQL=0.65% und σ .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Polya mit $\bar{p}$ und σ =0.3%

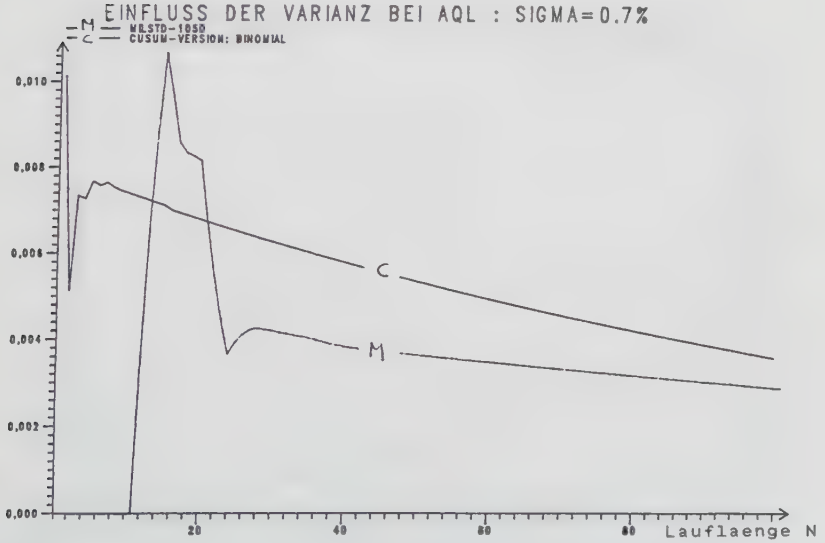

Abb. 8.4.b : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase unter einer Prozesskurve mit $\bar{p}$ =AQL=0.65% und σ .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Polya mit $\bar{p}$ und σ =0.7%

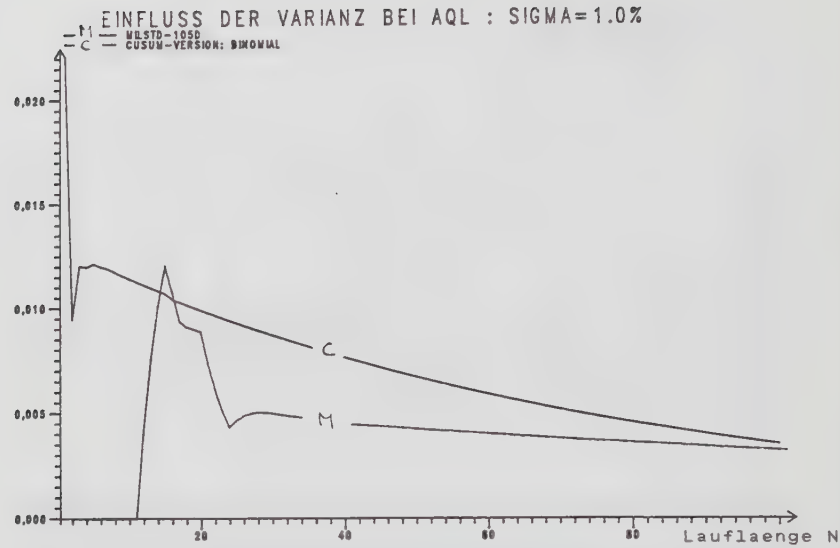

Abb. 8.4.c : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase unter einer Prozesskurve mit $\bar{p}$ =AQL=0.65% und σ .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Polya mit $\bar{p}$ und σ =1.0%

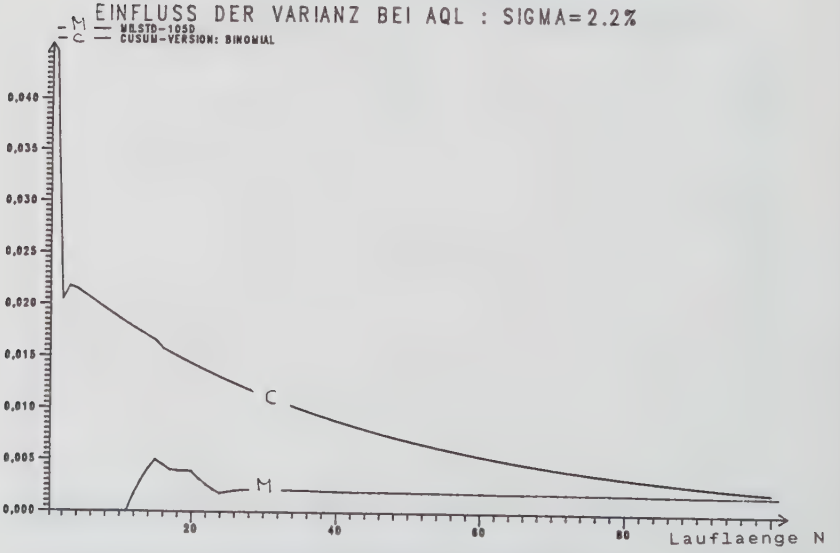

Abb. 8.4.d : Vergleich der Dichte P(N=n) der Lauflaenge N bis zum Erreichen Discontinue-Phase unter einer Prozesskurve mit $\bar{p}$ =AQL=0.65% und σ .
MIL-STD-105D : AQL=0.65% Code Letter K
Binomial-CUSUM-Version : h_T =3.1 h_D =4.73
Stichprobenverteilung : Polya mit $\bar{p}$ und σ =2.2%

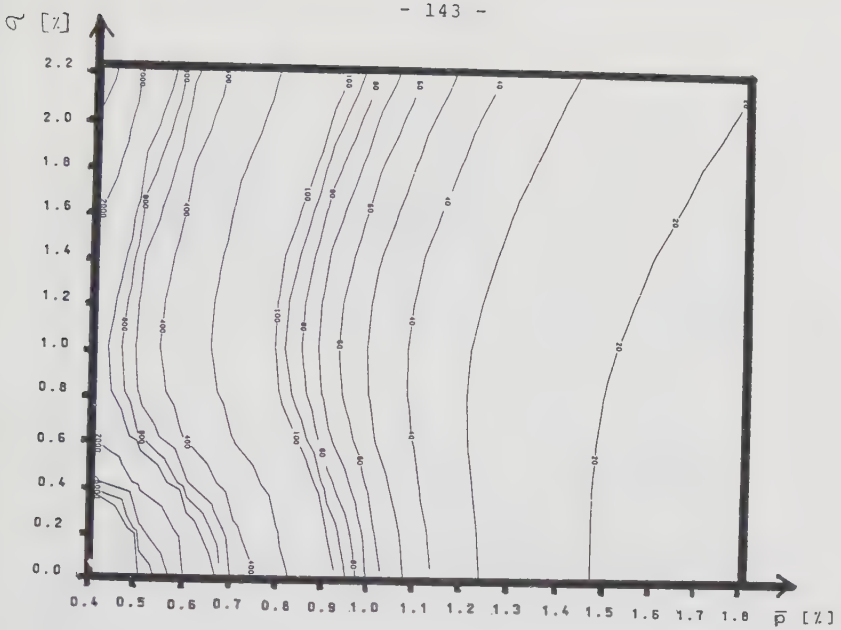

Abb. 8.5.a : ARL_D-Isolinien des MIL-STD-105D
AQL=0.65% CODE LETTER K
Stichprobenverteilung : Polya mit $\bar{p}$ und σ

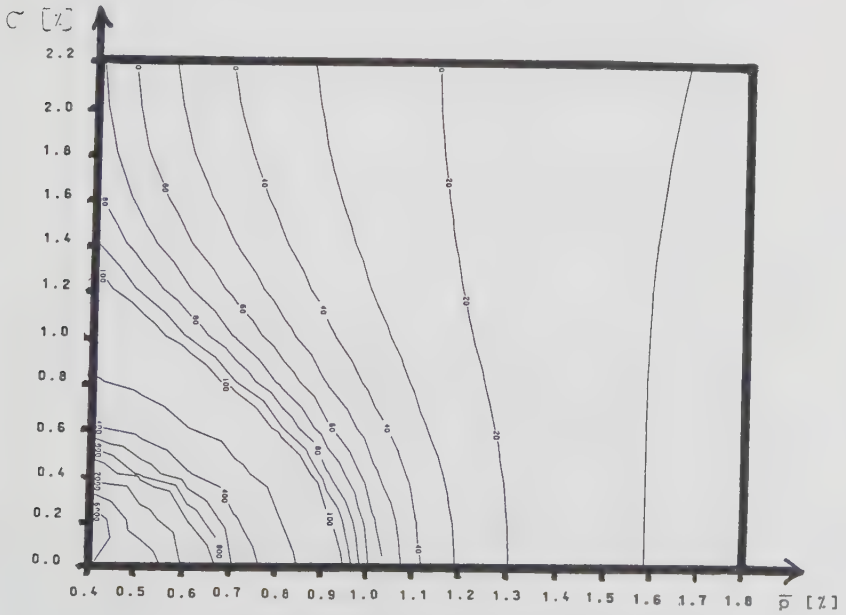

Abb. 8.5.b : ARL_D-Isolinien der Binomial-CUSUM-Version des MIL-STD-105D . h_T=3.1 h_p=4.73
Stichprobenverteilung : Polya mit $\bar{p}$ und σ

Die Situation ist ähnlich wie in Kapitel 6 bei den einfachen CUSUM-Karten zur Kontrolle der Prozeßkurve (vgl. Abb. 6.3 A1 und 6.3 A3 und Bemerkung 6.2). Für n = 125 werden Lose sowohl unter normaler als auch verschärfter Prüfung für $X_t \leq 1$ angenommen. Der unimodale Verlauf von ARL_D^{MIL} $(\bar{p} = AQL, \sigma)$ wird daher im wesentlichen durch den unimodalen Verlauf von $P_{\bar{p},\sigma}(X = 0) + P_{\bar{p},\sigma}(X = 1)$ bestimmt (vgl. Abb. 6.4 und Tab. 6.3).

In Abbildung 8.6 wird der Einfluß der Prozeßvarianz bei $\bar{p} = AQL = 0.65\ \%$ auf ARL_D von MIL-STD-105 D und den 4 in Abschnitt 8.2 genannten CUSUM-Versionen dargestellt. Es überrascht nicht, daß die CUSUM-Version über die Aufnahme-/Ablehnungs-Entscheidung (vgl. Definition von $GCU2_t$ über (8.5)) ein ähnlich unimodales Verhalten in ARL_D zeigt wie der MIL-STD-105 D. Analog zu der CUSUM-Karte zur Alternative (A1) in Kapitel 6 zeigt die CUSUM-Version des MIL-STD-105 D zu der gleichen Prozeßkurvenalternative (AMIL 1) $\bar{p}_A = 2.14\ \%$ und $\sigma_A = \sigma_S = 0.65\ \%$ ein unimodales Verhalten in ARL_D $(\bar{p} = AQL, \sigma)$. Das Verhalten der CUSUM-Version zu der Alternative (AMIL 2) $\bar{p}_A = 2.14\ \%$ und $\sigma_A = 2.14\ \%$ ist identisch mit der Binomial-CUSUM-Version.

Wir haben bisher den MIL-STD-105 D und seine CUSUM-Versionen nur unter der Annahme untersucht, daß die Stichprobenverteilung zeitlich konstant ist.

Wir betrachten jetzt den Fall, daß zu einem unbekannten Zeitpunkt τ $\bar{p}$ vom AQL-Wert 0.65 % auf $\bar{p}_A = 2.14\ \%$ springt. Die Prozeßvarianz σ^2 bleibe konstant. Es sei K_τ die Lauflänge nach dem Sprung:

$$K_\tau := \begin{cases} 0 & N \leq \tau \\ k & N = \tau + k \qquad k \in \mathbb{N} \end{cases} \tag{8.21}$$

und

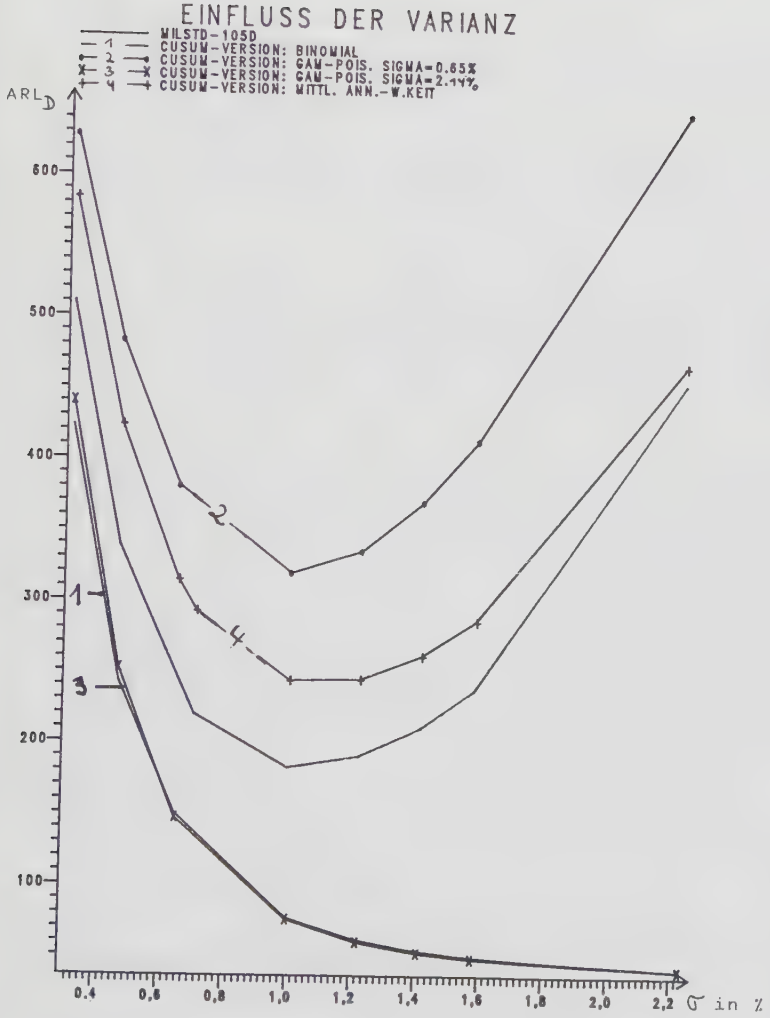

Abb. 8.6 : Der Einfluss der Prozessvarianz σ^2 auf ARL_D beim MIL-STD-105D und 4 verschiedenen CUSUM-Versionen des MIL-STD-105D bei $\bar{p}$ = AQL = 0.65% Stichprobenverteilung : Polya mit $\bar{p}$=0.65% und σ

(8.22) $ARL_{K_\tau}(E_z) := E(K_\tau | \text{Prüfsystem bei } t = \tau \text{ im Zustand } E_z)$

Da die Übergangsmatrix M_S vor dem Zeitpunkt τ die Voraussetzungen von Theorem 4.1 erfüllt, gilt nach Satz 4.14:

$$(8.23)\quad E(K_\tau | N > \tau) \xrightarrow[\tau \to \infty]{} \int ARL_{K_\tau}(E_z)\, m_S^\infty(E_z)\, \mu(dz)$$

wobei $m_S^\infty(E_z)$ gegeben ist durch:

$$(8.24)\quad m_S^\infty(E_z) := \lim_{\tau \to \infty} \frac{M_S^\tau(E_{Start}, E_z)}{\int M_S^\tau(E_{Start}, E_z)\, \mu(dz)}$$

Tabelle 8.5 zeigt die Werte von $ARL_{K_\tau}(E_z)$ und $m_S^\infty(E_z)$ für $\bar{p}_A$ = 1.25 AQL, 2 AQL und 2.0 % ≈ 3 AQL. Als Prozeßvarianz wurde $\sigma = 0$ gewählt, d. h. die Stichproben sind binomial-verteilt.

Die Planer des MIL-STD-105 D sind bei der Bestimmung der Kenngrößen von der impliziten Annahme ausgegangen, daß die bedingte mittlere Lauflänge von K_τ gut approximiert wird, wenn m_S^∞ durch eine 1-Punkt-Verteilung auf E_{Start} ersetzt wird, d. h.

$$(8.25)\quad E(K_\tau | N > \tau) \xrightarrow[\tau \to \infty]{} \int ARL_{K_\tau}(E_z)\, m_S^\infty(E_z)\, \mu(dz) \approx ARL_{K_\tau}(E_{Start})$$

Da sich die Werte $ARL_{K_\tau}(R)$ und $ARL_{K_\tau}(E_{Start})$ in Tabelle 8.5 nicht stark unterscheiden, dürfen wir annehmen, daß (8.25) im allgemeinen in guter Näherung gilt.

Zustand E_z	$ARL_{k_z}(E_z)$ p_A=0.8%	p_A=1.3%	p_A=2.0%	$m_S^{\infty}(E_z)$
R	237	32.8	18.3	0.635
9A	235	31.0	16.6	0.027
8A	233	29.6	15.7	0.028
7A	231	28.6	15.2	0.030
6A	229	27.8	14.9	0.031
5A	227	27.1	14.8	0.033
4A	226	26.6	14.7	0.035
3A	224	26.3	14.7	0.024
2A	223	26.0	14.6	0.025
A	222	25.7	14.6	0.027
START N	221	25.6	14.6	0.028
R	196	20.4	12.4	0.014
AR	202	21.4	12.6	0.013
AAR	209	22.8	12.9	0.012
AAAR	217	24.5	13.9	0.012
START T	106	12.3	10.0	0.002
T2	105	11.3	9.0	0.002
T3	104	10.3	8.0	0.002
T4	103	9.3	7.0	0.002
T5	102	8.3	6.0	0.002
T6	68.5	6.6	5.0	0.002
T7	55.5	5.3	4.0	0.002
T8	40.3	3.9	3.0	0.001
T9	22.4	2.4	2.0	0.001

Tabelle 8.5 : $ARL_{k_z}(E_z)$ und $m_S^{\infty}(E_z)$ fuer die transienten Pruefzustaende E_z des MIL-STD-105D . Verteilung vor dem Sprung : Binomial mit p_S = AQL = 0.65% nach dem Sprung : Binomial mit p_A =0.8%,1.3%,2.0%

Tabelle 8.6 vergleicht $ARL_{K_\tau}(E_{Start})$ und $ARL(K_\tau|N>\tau)$ für $\tau = 50$ für den MIL-STD-105 D und die Binomial-CUSUM-Version. Außerdem wird für die beiden Verfahren die Wahrscheinlichkeit für einen falschen Alarm $P(N \leq \tau)$ angegeben. Die in Tabelle 8.6 angegebene Prozeßsteuerung gelte vor und nach dem Sprung von $\bar{p}_S = 0.65$ % auf $\bar{p}_A = 0.8$ %, 1.3 %, 2.0 %. Die Ergebnisse bestätigen die Gültigkeit von (8.25).

Es ist bemerkenswert, daß der MIL-STD-105 D bei Lieferung von Losen mit AQL-Qualität in 6 % aller Fälle schon nach 50 Losen die Discontinue Phase erreicht hat. Dies zeigt, wie scharf die Prüfstufenregelung beim MIL-STD-105 D ist. Zu ähnlichen Ergebnissen gelangt auch Koyama (1974). Bei einer Prozeßvarianz von $\sigma = 1.0$ % bricht der MIL-STD-105 D in 23 % aller Fälle vor dem 50. Los die Warenannahme ab!

Der Vergleich von $E(K_\tau|N>\tau)$ beim MIL-STD-105 D und der Binominal-CUSUM-Version des MIL-STD-105 D zeigt die starke Überlegenheit der CUSUM-Version, Qualitätseinbrüche schnell aufzudecken: Bei der CUSUM-Version liegt $E(K_\tau|N>\tau)$ bei $\bar{p}_A = 2$ AQL um ca. 2/3 unter den entsprechenden Werten beim MIL-STD-105 D. Bei $\bar{p}_A = 2\ \% \approx 3$ AQL variiert dieses Verhältnis zwischen 30 % (bei $\sigma = 0.0$ %) und 43 % (bei $\sigma = 2.2$ %).

Parameter der Prozeßkurve nach Sprung		MIL-STD-105D			Binomial-CUSUM-Version des MIL STD-105D		
$\bar{p}_A$ in %	σ in %	$ARL_{K_\tau}(E_{START})$	$E[K_\tau \mid N>50]$	$P[N \leq 50]$	$ARL_{K_\tau}(E_{START})$	$E[K_\tau \mid N>50]$	$P[N \leq 50]$
	0.0	221	229	0.06	249	248	0.05
	0.3	162	167	0.10	161	158	0.10
	0.7	100	102	0.20	75.5	74.4	0.32
0.8	1.0	91.0	90.6	0.23	52.8	51.8	0.47
	1.2	93.4	92.5	0.22	44.2	44.5	0.56
	1.4	100	99.6	0.20	39.0	39.2	0.60
	1.6	110	109	0.18	37.3	37.7	0.64
	2.2	184	185	0.10	32.4	32.6	0.70
	0.0	25.6	30.3	0.06	19.7	19.4	0.05
	0.3	25.2	29.1	0.10	18.9	18.5	0.10
	0.7	25.0	26.8	0.20	17.5	17.2	0.32
1.3	1.0	25.6	26.6	0.23	16.6	16.2	0.47
	1.2	26.6	27.4	0.22	16.1	16.1	0.56
	1.4	27.8	28.6	0.20	15.8	15.7	0.60
	1.6	29.2	30.1	0.18	15.7	15.8	0.64
	2.2	37.4	39.3	0.10	15.7	16.3	0.70
	0.0	14.6	16.8	0.06	5.5	5.5	0.05
	0.3	14.8	16.5	0.10	5.5	5.6	0.10
	0.7	15.0	15.8	0.20	5.7	5.7	0.32
2.0	1.0	15.3	15.7	0.23	5.9	5.8	0.47
	1.2	15.6	15.8	0.22	6.0	6.1	0.56
	1.4	15.9	16.2	0.20	6.2	6.2	0.60
	1.6	16.2	16.7	0.18	6.4	6.6	0.64
	2.2	18.0	19.1	0.10	7.1	8.3	0.70

Tabelle 8.6 : Vergleich von $ARL_{K_\tau}[E_{START}]$, $E[K_\tau \mid N>50]$ und $P[N \leq 50]$ beim MIL-STD-105D und der Binomial-CUSUM-Version des MIL-STD-105D
Stichprobenverteilung vor dem Sprung : Polya mit $\bar{p}_s$=AQL=0.65% und σ , nach dem Sprung : Polya mit $\bar{p}_A$=0.8%,1.3%,2.0% und σ .

Abschnitt 8.6:

Vergleich von Prüfaufwand und Durchschlupf beim MIL-STD-105 D und der Binomial-CUSUM-Version

Wir wollen jetzt den Prüfaufwand und den Durchschlupf für den MIL-STD-105 D und die Binomial-CUSUM-Versionen untersuchen, da diese beiden Kriterien letztendlich bei der Beurteilung der betrachteten Prüfsysteme ausschlaggebend sind.

Es sei (n_t, c_t) der Stichprobenplan, mit dem das t. Los geprüft wird. Definiere den Stichprobenumfang SN_t $(t \geq 1)$ durch:

$$SN_t := \begin{cases} n_t & N \geq t \\ o & N < t \end{cases}$$

Hinsichtlich der Lose wollen wir annehmen, daß abgelehnte Lose aussortiert werden und daß bei der Stichprobe schlechte Stücke durch gute Stücke ersetzt werden (Rectifying Inspection). Damit erhalten wir für den Durchschlupf (Outgoing Quality):

$$OQ_t := \begin{cases} 0 & X_t > c_t \quad \text{und} \quad N \geq t \\ \dfrac{Y_t}{N_L} & X_t \leq c_t \quad \text{und} \quad N \geq t \\ 0 & N < t \end{cases}$$

Hierbei ist Y_t die Anzahl der schlechten Stücke im Losrest.

Satz 8.3:

Ist die Stichprobenverteilung von X_t durch eine Binomial-Verteilung mit p und $n_t = n$ gegeben, so gilt für ein Los vom Umfang N_L:

$$(8.26)\quad E\left(0Q_t \mid (n_t, c_t) = (n, c)\right) = \left(1 - \frac{n}{N_L}\right) p \sum_{x=0}^{c} b(x \mid n, p)$$

Ist die Stichprobenverteilung von X_t durch eine Polya-Verteilung mit α, β und $n_t = n$ gegeben, so gilt:

$$(8.27)\quad E\left(0Q_t \mid (n_t, c_t) = (n, c)\right) = \left(1 - \frac{n}{N_L}\right) \sum_{x=0}^{c} \frac{\alpha + x}{\alpha + \beta + n} f(x \mid n, \alpha, \beta)$$

Hierbei ist $f(x|n, \alpha, \beta)$ durch (6.5) gegeben.

Beweis:

Analog zum Beweis von Satz 6.1 erhält man:

$$E\left(0Q_t \mid (n_t, c_t) = (n, c)\right) = \frac{\Lambda}{N_L} \sum_{X=0}^{N_L} \left(\sum_{x=0}^{c} (X - x) h(x \mid n, X, N_L) \right) P(X_L = X)$$

$$= \frac{\Lambda}{N_L} \sum_{y=0}^{N_L - n} \sum_{x=0}^{c} \int_0^{\Lambda} y \, b(y \mid N_L - n, p) \, b(x \mid n, p) \cdot w(p) \, dp$$

$$= \frac{\Lambda}{N_L} \sum_{x=0}^{c} \int_0^{\Lambda} (N_L - n) p \, b(x \mid n, p) \, w(p) \, dp$$

Für $\sigma = 0$ folgt hieraus sofort (8.26). Für $\sigma > 0$ (d. h. im Polya-Fall) erhält man:

$$\int_0^1 p\, b(x|n,p)\, w(p)\, dp = \binom{n}{x} \frac{Be(x+\alpha+1, n-x+\beta)}{Be(\alpha,\beta)}$$

$$= \binom{n}{x} \frac{x+\alpha}{\alpha+\beta+n} \frac{(x+\alpha-1)\cdots\alpha\,(n-x+\beta-1)\cdots\beta}{(n+\alpha+\beta-1)\cdots(\alpha+\beta)}$$

$$= \frac{\alpha+x}{\alpha+\beta+n}\, f(x|n,\alpha,\beta)$$

□

Es bezeichne S_t die Prüfstufe auf der das t. Los geprüft wird. Es sei Z_s $s \in \{R, N, T\}$ diejenige Teilmenge der transienten Prüfzustände, die zur Prüfstufe s gehören. Man erhält dann:

$$P(S_t = \rho) = \begin{cases} 0 & N < t \\ \int_{Z_\rho} M_S^{t-1}(E_{Start}, E_z)\, \mu(dz) & t \le \tau \le N \\ \int_{Z_\rho} \left(\int M_S^{\tau}(E_{Start}, E_y)\, M_A^{t-\tau-1}(E_y, E_z)\, \mu(dy) \right) \mu(dz) & \tau < t \le N \end{cases}$$

Für uns von Interesse sind dann der erwartete Prüfaufwand zur Zeit t:

$$ASN_t := E(SN_t) = \sum_{\rho \in \{R,N,T\}} n_\rho\, P(S_t = \rho)$$

und der erwartete Durchschlupf (Average Outgoing Quality) zur Zeit t:

$$AOQ_t := E(OQ_t) = \sum_{\rho \in \{R,N,T\}} E\left(OQ_t \mid (n_t, c_t) = (n_\rho, c_\rho)\right) P(S_t = \rho)$$

Wenn kein Sprung vom Sollzustand in den Alarmzustand stattfindet, werden hohe Lauflängen erwartet. Es ist daher nicht sinnvoll, den absoluten Prüfaufwand und Durchschlupf miteinander zu vergleichen, sondern:

$$CASN_t := \frac{E(SN_t)}{P(N \geq t)} \quad \text{und} \quad CAOQ_t := \frac{E(OQ_t)}{P(N \geq t)} \qquad t \geq 1$$

Analog zu Satz 4.14 läßt sich zeigen, daß $P(S_t = s | N \geq t)$, $CASN_t$ und $CAOQ_t$ für $t \to \infty$ konvergieren. Bezeichne diese Grenzwerte mit $CP_\infty(s)$, $CASN_\infty$ und $CAOQ_\infty$. Es gilt (vgl. (8.23) und (8.24)):

$$(8.28) \qquad CP_\infty(s) = \int_{Z_s} m_s^\infty(E_z)\mu(dz) \qquad s \in \{R,N,T\}$$

$$(8.29) \qquad CASN_\infty = \sum_{s \in \{R,N,T\}} n_s CP_\infty(s)$$

$$(8.30) \qquad CAOQ_\infty = \sum_{s \in \{R,N,T\}} E(OQ_t | (n_t,c_t) = (n_s,c_s))\, CP_\infty(s)$$

Tabelle 8.7 vergleicht $CASN_\infty$, $CAOQ_\infty$ und $CP_\infty(s)$, s = R,N,T für den MIL-STD-105 D und die Binomial-CUSUM-Version für den Fall σ = 0 und p = AQL = 0.65 % (d. h. Stichprobenverteilung = Binomial-Verteilung).

	MIL-STD-105D	Binomial-CUSUM Version des MIL-STD-105D
$CASN_\infty$	77.3	73.3
$CAOQ_\infty$	0.61%	0.61%
$CP_\infty[R]$	0.635	0.688
$CP_\infty[N]$	0.345	0.303
$CP_\infty[T]$	0.020	0.009

Tabelle 8.7 : Vergleich des asymptotischen Pruefaufwands $CASN_\infty$ und asymptotischen Durchschlupfs $CAOQ_\infty$ sowie der asymptotischen Pruefstufenwahrscheinlichkeiten $CP_\infty[R]$, $CP_\infty[N]$, $CP_\infty[T]$ beim MIL STD 105D und der Binomial CUSUM-Version des MIL-STD-105D . Stichprobenverteilung : Binomial mit p=AQL=0.65% .

Obwohl beide Verfahren zunächst nur hinsichtlich ihrer Lauflänge angepaßt wurden, stimmen sie auch hinsichtlich ihres Prüfaufwands und Durchschlupfs für p = AQL gut überein.

Für die Zeit nach dem Sprung von $\bar{p}_S$ nach $\bar{p}_A$ erwarten wir eine kurze Laufzeit. In diesem Fall ist es sinnvoll, den absoluten Prüfaufwand und den absoluten Durchschlupf nach dem Sprung von $\bar{p}_S$ auf $\bar{p}_A$ zu vergleichen.

Der durchschnittliche Prüfaufwand und der durchschnittliche Durchschlupf über eine Periode von K Losen nach dem Sprung von $\bar{p}_S$ auf $\bar{p}_A$ ist gegeben durch:

$$\overline{ASN}(K,\tau) := \frac{1}{K} \sum_{k=1}^{K} \frac{ASN_{k+\tau}}{P(N>\tau)} \tag{8.31}$$

$$\overline{AOQ}(K,\tau) := \frac{1}{K} \sum_{k=1}^{K} \frac{AOQ_{k+\tau}}{P(N>\tau)} \tag{8.32}$$

Die Abhängigkeit von τ in (8.31) und (8.32) läßt sich beseitigen, indem man $\tau > 30$ annimmt, und die bedingte Dichte auf den transienten Prüfzuständen durch ihren Grenzwert m_S^{∞} (vgl. (8.24)) ersetzt.

Die Abhängigkeit von K in (8.31) und (8.32) ist inhaltlicher Natur. Für $K \to \infty$ streben $\overline{ASN}(K,\tau)$ und $\overline{AOQ}(K,\tau)$ gegen Null.

Von Interesse ist auch der durchschnittliche Prüfaufwand und Durchschlupf bei den noch geprüften Losen, nachdem ein Sprung von $\bar{p}_S$ nach $\bar{p}_A$ eingetreten ist:

(8.33) $$\overline{CASN}(k,\tau) := \frac{1}{k}\sum_{k=1}^{k}\frac{ASN_{k+\tau}}{P(N\geq k+\tau)}$$

(8.34) $$\overline{CAOQ}(k,\tau) := \frac{1}{k}\sum_{k=1}^{k}\frac{AOQ_{k+\tau}}{P(N\geq k+\tau)}$$

Die bedingte Verteilung auf den transienten Prüfzuständen gegeben $N \geq k+\tau$ konvergiert für $k\to\infty$ gegen (vgl. Satz 4.12):

(8.35) $$m_A^{\infty}(E_z) := \frac{\psi_A(E_z)}{\int \psi_A(E_z)\,\mu(dz)}$$

Damit konvergiert $P(S_{k+\tau} = s \mid N \geq k+\tau)$ für $k\to\infty$ gegen einen Grenzwert $CP_{\infty,\tau}(s)$:

(8.36) $$CP_{\infty,\tau}(\sigma) = \int_{Z_\sigma} m_A^{\infty}(E_z)\,\mu(dz) \qquad \sigma\in\{R,N,T\}$$

$\frac{ASN_{k+\tau}}{P(N\geq k+\tau)}$ und $\frac{AOQ_{k+\tau}}{P(N\geq k+\tau)}$ konvergieren für $k\to\infty$ damit gegen die Grenzwerte $CASN_{\infty,\tau}$ und $CAOQ_{\infty,\tau}$ wobei gilt:

(8.37) $$CASN_{\infty,\tau} = \sum_{\sigma\in\{R,N,T\}} n_\sigma\, CP_{\infty,\tau}(\sigma)$$

(8.38) $$CAOQ_{\infty,\tau} = \sum_{\sigma\in\{R,N,T\}} E\left(OQ_t \mid (n_t,c_t) = (n_\sigma,c_\sigma)\right) CP_{\infty,\tau}(\sigma)$$

Da die Summanden in (8.33) und (8.34) gegen $CASN_{\infty,\tau}$ und $CAOQ_{\infty,\tau}$ konvergieren, konvergiert auch das Cesaro-Mittel gegen diesen Grenzwert:

(8.39) $\overline{CASN}(K,\tau) \xrightarrow[K\to\infty]{} CASN_{\infty,\tau}$

(8.40) $\overline{CAOQ}(K,\tau) \xrightarrow[K\to\infty]{} CAOQ_{\infty,\tau}$

Tabelle 8.8 vergleicht $\overline{ASN}(K = 50, \tau > 30)$ $\overline{AOQ}(K = 50, \tau > 30)$ beim MIL-STD-105 D und der Binomial-CUSUM-Version. Als Stichprobenverteilung wurde die Binomialverteilung gewählt. $\overline{p}_A = p$ variiert in Tabelle 8.8 von 0.7 % bis 1.8 %. Zusätzlich sind in Tabelle 8.8 die Werte von $\overline{CASN}(K = 50, \tau > 30)$, $\overline{CAOQ}(K = 50, \tau > 30)$, sowie $CASN_{\infty,\tau}$ $CAOQ_{\infty,\tau}$ und $CP_{\infty,\tau}(s)$, $s \in \{R,N,T\}$ aufgelistet.

Die kürzere Laufzeit der CUSUM-Version nach dem Sprung drückt sich dabei in kleineren Werten von $P(K_\tau > 50 | N > \tau)$ aus.

Der gemittelte Prüfaufwand $\overline{ASN}$ liegt bei der CUSUM-Version deutlich unter dem des MIL-STD-105 D: Bei $\overline{p}_A = 1.3\ \% = 2$ AQL beträgt der Prüfaufwand der CUSUM-Version nur noch 62.3 % des Prüfaufwands des MIL-STD-105 D.

Dieser niedrige Prüfaufwand wird dadurch erreicht, daß die CUSUM-Version erheblich mehr Lose unter reduzierter Prüfung prüft: Beim MIL-STD-105 D werden z. B. bei $\overline{p}_A = 1.3\ \%$ 14 % aller geprüften Lose mit reduzierter Prüfung geprüft, bei der CUSUM-Version sind es jedoch 31 %. Trotz dieses verringerten Prüfaufwands liegt der absolute Durchschlupf $\overline{AOQ}$ der CUSUM-Version gleichmäßig unter den entsprechenden Werten des MIL-STD-105 D. Die Differenz der Maxima des Durchschlupfs bei den geprüften Losen beträgt lediglich 0.15 % zu ungunsten des CUSUM-Verfahrens.

MIL-STD-105D

Werte fuer p_A in % nach dem Sprung

	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
$P[K_\tau>50 \mid N>\tau]$	0.92	0.83	0.69	0.52	0.35	0.21	0.11	0.05	0.02	0.009	0.003	0.001
$\widehat{ASN}(K=50,\tau>30)$	77.4	80.0	80.0	77.0	71.6	65.0	58.5	52.6	47.9	44.2	41.3	39.1
$\widehat{AOQ}(K=50,\tau>30)$ in %	0.63	0.67	0.68	0.66	0.62	0.56	0.50	0.45	0.40	0.36	0.33	0.31
$\widehat{CASN}(K=50,\tau>30)$	80.3	86.4	92.3	97.7	102.5	106.6	110.0	112.7	114.9	116.6	117.9	118.9
$\widehat{CAOQ}(K=50,\tau>30)$ in %	0.65	0.72	0.78	0.83	0.86	0.88	0.88	0.87	0.84	0.82	0.78	0.75
$CASN_{\infty,\tau}$	80.8	87.8	94.6	100.7	106.1	110.6	114.2	117.0	119.2	120.8	122.0	122.8
$CAOQ_{\infty,\tau}$ in %	0.65	0.72	0.78	0.82	0.85	0.85	0.84	0.82	0.79	0.75	0.71	0.67
$CP_{\infty,\tau}$ [R]	0.59	0.49	0.40	0.32	0.25	0.19	0.14	0.10	0.07	0.05	0.04	0.03
$CP_{\infty,\tau}$ [N]	0.38	0.45	0.49	0.50	0.49	0.46	0.41	0.36	0.30	0.25	0.19	0.14
$CP_{\infty,\tau}$ [T]	0.03	0.06	0.11	0.18	0.26	0.35	0.45	0.54	0.63	0.70	0.77	0.83

Binomial-CUSUM-Version des MIL-STD-105D

Werte fuer p_A in % nach dem Sprung

	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
$P[K_\tau>50 \mid N>\tau]$	0.92	0.82	0.67	0.48	0.29	0.14	0.06	0.02	0.006	0.001	<0.001	<0.001
$\widehat{ASN}(K=50,\tau>30)$	72.5	72.8	70.0	63.8	55.0	45.3	36.5	29.3	23.9	20.0	17.1	14.9
$\widehat{AOQ}(K=50,\tau>30)$ in %	0.63	0.67	0.67	0.63	0.55	0.46	0.37	0.30	0.25	0.21	0.18	0.16
$\widehat{CASN}(K=50,\tau>30)$	75.5	79.9	84.3	88.7	92.9	96.8	101.3	103.1	107.1	108.0	109.9	111.5
$\widehat{CAOQ}(K=50,\tau>30)$ in %	0.66	0.73	0.81	0.87	0.92	0.97	1.00	1.02	1.03	1.03	1.02	1.01
$CASN_{\infty,\tau}$	75.5	80.1	84.8	89.4	93.8	97.8	101.3	104.4	107.1	109.4	111.4	113.0
$CAOQ_{\infty,\tau}$ in %	0.66	0.74	0.81	0.87	0.93	0.97	1.00	1.01	1.02	1.02	1.01	1.00
$CP_{\infty,\tau}$ [R]	0.66	0.59	0.53	0.47	0.41	0.36	0.31	0.27	0.23	0.20	0.18	0.16
$CP_{\infty,\tau}$ [N]	0.33	0.39	0.43	0.48	0.51	0.54	0.56	0.57	0.58	0.58	0.57	0.56
$CP_{\infty,\tau}$ [T]	0.01	0.02	0.04	0.05	0.08	0.10	0.13	0.16	0.19	0.22	0.25	0.28

Tabelle 8.8 : Vergleich von Durchschlupf und Pruefaufwand beim MIL-STD-105D und der Binomial-CUSUM-Version des MIL-STD-105D . Vergleich fuer die ersten 50 Lose nach einem Sprung von p_S=AQL=0.65% auf p_A=0.7%,...,1.8% .Stichprobenverteilung:Binomial.

Wir wollen nun noch den Einfluß der Prozeßvarianz bei $\bar{p}$ = AQL = 0.65 % und stationärer Verteilungsannahme untersuchen. σ variiert dabei zwischen 0.0 % und 2.2 %.

Tabelle 8.9 vergleicht sowohl die langfristigen bedingten Werte $CASN_\infty$, $CAOQ_\infty$ und $CP_\infty(s)$ $s \in \{R,N,T\}$ als auch die kurzfristigen Durchschnittswerte $\overline{ASN}(50) = \frac{1}{50} \sum_{t=1}^{50} ASN_t$ $\overline{AOQ}(50) = \frac{1}{50} \sum_{t=1}^{50} AOQ_t$.

Es fällt auf, daß bei größeren Prozeßvarianzen die CUSUM-Version bei den geprüften Losen mit 2/3 des Prüfaufwandes des MIL-STD-105 D immer noch AOQ-Werte weit unterhalb der AQL-Grenze liefert. Der absolute Durchschlupf $\overline{AOQ}$ innerhalb der ersten 50 Lose ist für beide Verfahren annähernd gleich. Der mittlere Prüfaufwand ist jedoch sehr unterschiedlich. Bei σ = 2.2 % beträgt der mittlere Prüfaufwand der CUSUM-Version nur noch 38 % des Prüfaufwands des MIL-STD-105 D.

Bemerkung 8.4:

Grob gesprochen, kann das unterschiedliche Verhalten von MIL-STD-105 D und der CUSUM-Versionen so beschrieben werden: Während die CUSUM-Version die meisten Lose auf reduzierter Prüfung prüft und relativ schnell Prozeßverschlechterungen erkennt, die zum Abbruch der Warenannahme führen, nimmt der MIL-STD-105 D für eine längere Zeit noch Lose an, die allerdings verschärft geprüft werden.

Ändert der Anlieferer seine Prozeßkurve nicht, so wird der größte Teil der Warenlose zurückgewiesen. Das heißt, ein Großteil des Prüfaufwands liegt darin, Warenlose zurückzuweisen, während die CUSUM-Version in dieser Situation die Warenannahme abbricht und den Anlieferer zu einer Verbesserung seiner Produktion zwingt.

□

MIL-STD-105D

Werte fuer die Prozessstreuung σ in %

	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2
P[N> 50]	0.95	0.92	0.87	0.82	0.78	0.76	0.77	0.79	0.82	0.84	0.87	0.89
$\overline{ASN}$(50)	85.0	86.1	87.8	89.3	90.4	91.4	92.5	93.4	94.2	94.6	94.8	94.6
$\overline{AOQ}$(50) in %	0.58	0.57	0.53	0.47	0.41	0.35	0.31	0.27	0.24	0.21	0.19	0.17
$CASN_\infty$	77.3	79.3	83.5	88.3	92.1	94.5	95.6	95.7	95.2	94.2	93.0	91.6
$CAOQ_\infty$ in %	0.60	0.60	0.58	0.53	0.48	0.42	0.37	0.32	0.28	0.25	0.22	0.20
CP_∞[R]	0.64	0.61	0.55	0.49	0.44	0.41	0.39	0.39	0.40	0.41	0.43	0.45
CP_∞[N]	0.34	0.37	0.40	0.44	0.47	0.49	0.50	0.51	0.51	0.50	0.50	0.49
CP_∞[T]	0.02	0.02	0.05	0.07	0.09	0.10	0.11	0.10	0.09	0.09	0.07	0.06

Binomial-CUSUM-Version des MIL-STD-105D

Werte fuer die Prozessstreuung σ in %

	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2
P[N> 50]	0.95	0.92	0.85	0.74	0.62	0.52	0.45	0.40	0.36	0.33	0.31	0.30
$\overline{ASN}$(50)	72.4	71.5	68.5	63.3	57.3	51.9	47.4	43.9	41.2	39.1	37.5	36.3
$\overline{AOQ}$(50) in %	0.59	0.58	0.55	0.49	0.42	0.36	0.30	0.26	0.23	0.20	0.17	0.15
$CASN_\infty$	73.3	73.3	72.9	72.1	70.7	69.1	67.5	65.8	64.3	62.9	61.6	60.5
$CAOQ_\infty$ in %	0.61	0.60	0.59	0.56	0.53	0.49	0.45	0.40	0.41	0.33	0.30	0.27
CP_∞[R]	0.69	0.69	0.69	0.70	0.72	0.74	0.77	0.79	0.81	0.83	0.84	0.86
CP_∞[N]	0.30	0.30	0.29	0.28	0.26	0.24	0.21	0.19	0.17	0.16	0.14	0.13
CP_∞[T]	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01

Tabelle 8.9 : Einfluss der Prozessstreuung σ auf den Pruefaufwand und den Durchschlupf von MIL-STD-105D und der Binomial-CUSUM-Version des MIL-STD-105D Stichprobenverteilung : Polya mit $\bar{p}$ =AQL=0.65% und σ .

Abschnitt 8.6:

Abschließende Bemerkungen

Bemerkung 8.5:

Der MIL-STD-105 D wurde hier ohne die Regelung betrachtet, nach der die Gesamtanzahl schlechter Stücke in den 10 letzten Stichproben nicht größer als eine Zahl c_{10} sein darf, um von normaler Prüfung in die reduzierte Prüfung zu gelangen. Die Beachtung dieser Regel führt zu kürzeren Laufzeiten bei $\bar{p}_S$ = AQL und zu einem vergrößerten Prozentsatz von Losen unter normaler Prüfung.

Für $n_N * AQL = 125 * 0.65\ \% \approx 0.8$ findet man bei Hald (1981), p. 96, die Werte:

$$ARL_D = 693 \text{ und}$$

$$CP_{\cdot}(R) = 0.317,\ CP_{\cdot}(N) = 0.64\ CP_{\cdot}(T) = 0.035.$$

Der Prüfaufwand bei den geprüften Losen $CASN_{\infty}$ erhöht sich damit (vgl. (8.29)) von 77.3 auf 101.2. Der bedingte Durchschlupf bei $CASN_{\infty} = 77.3$ beträgt aber $CAOQ_{\infty} = 0.6.\%$ und liegt damit unterhalb der AQL-Grenze, so daß die Steigerung des Prüfaufwands unter dem AOQ-Kriterium nicht gerechtfertigt erscheint.

Der MIL-STD-105 D ohne diese Zusatzregel hält auch für $\bar{p}_A > AQL$ diese AOQ-Obergrenze gut ein (vgl. $\overline{ASN}(K=50, \tau>30)$ in Tabelle 8.8), so daß diese Zusatzregel, abgesehen von ihrer Unhandlichkeit, einen ungerechtfertigten Mehraufwand bei der Stichprobenprüfung erzeugt.

□

Bemerkung 8.6:

Die CUSUM-Versionen lassen sich leicht auf mehr als drei Prüfstufen verallgemeinern.

□

Bemerkung 8.7:

Eine Übertragung der CUSUM-Prüfstufensteuerung auf den Fall der Variablenprüfung (MIL-STD-414) liegt auf der Hand. Da die Prüfstufensteuerung sehr ähnlich der des MIL-STD-105 D ist, darf man daher ähnliche Resultate hinsichtlich des Verhältnisses von AOQ und ASN erwarten.

Schneider und Wilrich (1984) kommen in einer Simulationsstudie über eine Prüfstufensteuerung mit 2 CUSUM-Karten zu ähnlichen Ergebnissen wie in den Abschnitten 8.4 und 8.5. □

Bemerkung 8.8:

Die CUSUM-Prüfstufenregelung basierte auf den Zuwachsfunktionen

$$Y_{(\sigma)}(x) = \ln \frac{f(x \mid n_\sigma, \bar{p}_A, \sigma_A)}{f(x \mid n_\sigma, \bar{p}_S, \sigma_S)} \qquad \sigma \in \{R, N, T\}$$
$$x = 0, 1, \ldots, n_\sigma$$

In der Praxis bereitet die Angabe von Prozeßkurven mit genauer Festlegung von $\bar{p}$ und σ häufig Schwierigkeiten, da nicht genügend Informationen über $\bar{p}_S$, $\bar{p}_A$, σ_S und σ_A vorliegen.

Es ist jedoch denkbar, die Y(s) $s \in \{R,N,T\}$ direkt anzugeben und damit auf eine Angabe von $\bar{p}_S$, $\bar{p}_A$, σ_S und σ_A zu verzichten.

Die Angabe von Y(s) (x) drückt dann eine subjektive Bewertung des Stichprobenergebnisses x bei Prüfstufe s in Richtung Verlängerung (Y(s) (x) < 0) oder in Richtung Abbruch der Warenannahme (Y(s) (x) > 0) aus. □

Literaturverzeichnis

Abel, V. und Fuchs, E. (1984): "On the Run Length of One-Sided CUSUM-Procedures in the Change Point Case", erscheint in Proceedings 9. Symposium über OR, August 84

Barnard, G. A. (1959): "Control charts and stochastic processes", Journal of the Royal Statistical Society, Ser. B, 21, 239 - 257

Bauer, P. und Hackl, P. (1978): "The use of MOSUMs for quality control", Technometrics, 20, 431 - 436

Bauer, P. und Hackl, P. (1982): "A parabolic mask fc continous sampling inspection", Z. Oper. Res., 26, B 47 - B 54

Bissell, A. F. (1969): "CUSUM-Techniques for Quality Control", Applied Statistics, 18, 1 - 30

Bissell, A. F. (1979): "A semi-parabolic mask for CUSUM charts", The Statistician, 28, 1 - 7

Booth, N. B. and Smith, A. F. M. (1976): "Batch Acceptance Schemes Based on an Autoregressive Prior", Biometrika, 63, 133 - 136

Brook, D. and Evans, D. A. (1972): "An Approach to the Probability Distribution of CUSUM Rum Length", Biometrika, 59, 539 - 549

Brown, G. G. and Rutemiller, H. C. (1973): "A Cost Analysis of Sampling Inspection under Military Standard 105 D", Naval Research Logistics Quarterly, 20, 181 - 199

Bruyn, v. D. d. (1968): "Cumulative Sum Test: Theory and Practice", Griffin, London

Campling, G. E. G. (1968): "Serial Sampling Acceptance Schemes for Large Batches of Items where the Mean Quality has a Normal Prior Distribution", Biometrika, 55, 393 - 399

Case, K. E. and Keats, J. B. (1982): "On the Selection of a Prior Distribution in Bayesian Acceptance Sampling", Journal of Quality Technology, 14, 10 - 17

Cox, D. R. (1960): "Serial Sampling Acceptance Schemes Derived from Bayes' Theorems", Technometrics, 2, 353 - 360

Cox, D. R. and Miller, H. D. (1965): "The Theory of Stochastic Processes", Methuen, London

Ewan, W. D. and Kemp, K. W. (1960): "Sampling Inspection of Continous Processes with No Autocorrelation Between Successive Results", Biometrika, 47, 363 - 380

Hald, A. (1960): "The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs", Technometrics, 2, 275 - 340

Hald, A. (1981): "Statistic Theory of Sampling Inspection by Attributes", Academic Press, London

Harris, T. E. (1963): "The Theory of Branching Processes", Springer Verlag, Berlin

Johnson, N. L. (1961): "A simple theoretical approach to cumulative sum charts", J. Amer. Statist. Assoc., 56, 835 - 840

Johnson, N. L. and Leone, F. C. (1962 a): "Cumulative Sum Control Charts. Mathematical principles applied to their construction and use", Part I, Industrial Quality Control, 18, 12, 15 - 21

Johnson, N. L. and Leone, F. C. (1962 b): "Cumulative sum control charts. Mathematical principles applied to their construction and use", Part II, Industrial Quality Control, 19, 1, 29 - 36

Khan, R. A. (1978): "Wald's approximation to the average run length in CUSUM Procedures", J. Statistical Planning and Inference, 2, 63 - 77

Khan, R. A. (1981): "A note on Page's two-sided cumulative sum procedure", Biometrika, 68, 717 - 719

Khan, R. A. (1984): "On cumulative sum procedures and the SPRT applications", J. Roy. Statist. Soc., Ser. B, 46, 79 - 85

Koyama, T. (1974): "Analysis on the Dynamic Characteristics of Severity Control in Sampling Inspection Plans with Severity Adjustment", PH. D. Thesis, Keio University, Tokyo

Koyama, T. (1978): "Modified Switching Rules for Sampling Schemes such as MIL-STD-105 D", Technometrics, 20, 95 - 102

Lai, T. L. (1974): "Control Charts based on weightened Sums", Ann. Statist., 2, 134 - 147

Lenz, H. - J. und Rendtel, U. (1984): "Performance Evaluation of the MIL-STD-105 D, Skip-Lot Sampling Plans, and BAYESIAN Single Sampling Plans", in "Frontiers in Statistical Quality Control 2", Lenz, Wetherill, Wilrich eds, Physica Verlag, Würzburg

Lucas, R. W. and Crosier, R. B. (1982): "Fast initial Response for CUSUM-Quality Control Schemes", Technometrics, 24, 199 - 205

MIL-STD-105 D (1963): "Sampling Procedures and Tables for Inspection by Attributes", Department of Defense, Washington, D. C.

Mood, A. M. (1943): "On the dependence of sampling inspection plans upon population distributions", Ann. Math. Statist., 14, 415 - 425

Page, E. S. (1954): "Continuous inspection Schemes", Biometrika, 41, 100 - 115

Reimann, A. (1984): "Kostenoptimale Inspektionsstrategien für den Fall zweier stochastisch abhängiger Losschlechtanteile", in "Arbeiten zur Angewandten Statistik", Schäffer, Schönfeld, Wetzel Hrg, Physica Verlag, Würzburg

Reetz, D. (1984): "Optimal Skip Lot Single Sampling Plans for Markov Chains", in "Frontiers in Statistical QuaLITY Control 2", Lenz, Wetherill, Wilrich eds, Physica Verlag, Würzburg

Reynolds, M. R. (1975): "Approximations to the average run legth in Cumulative Sum control Charts", Technometrics, 17, 65 - 71

Roberts, S. W. (1966): "A comparison of some control chart procedures", Technometrics, 8, 411 - 430

Schneider, H. und Wilrich, P. (1984): "A New System of Switching Rules for ISO 3951 using CUSUM Methods", in "Frontiers in Statistical Quality Control 2", Lenz, Wetherill, Wilrich eds, Physica Verlag, Würzburg

Waldmann, K. - W. (1983): "Bounds to the average run length in CUSUM quality-control schemes", erscheint in Technometrics

Waldmann, K. - H. (1984 a): "Bounds to the distribution of the run length of one-sided CUSUM quality control schemes", erscheint in Technometrics

Waldmann, K. - H. (1984 b): "Bounds to the distribution of the run length in general quality control schemes", zur Veröffentlichung eingereicht

Woodall, W. H. (1983): "The distribution of the run length of one-sided CUSUM-Procedures for Continuous Random Variables", Technometrics, 25, 295 - 301

Woodall, W. H. (1984): "On the Markov Chain Approach to the Two-Sided CUSUM Procedure", Technometrics, 26, 41 - 46

Zacks, S. (1981): "The Probability Distribution and the Expected Value of a Stopping Variable Associated With One-Sided CUSUM Procedures for Non-Negative Integer Valued Random Variables", Communications in Statistics, A10, 2245 - 2258

Lebenslauf

Name : Ulrich Rendtel

Geburtsdatum : 18.10.1950

Geburtsort : Fuerstenberg/Havel

1957	: Einschulung in Duesseldorf
1961-1970	: Besuch des Lessing-Gymnasiums in Duesseldorf
16.6.1970	: Abitur
Okt. 70 - Feb. 71	: Wehrdienst
Feb. 71 - April 72	: Zivildienst
WS 71/72 - SS 72	: Mathematikstudium an der Universitaet Duesseldorf
WS 72/73 - SS 79	: Mathematikstudium an der Freien Universitaet Berlin Studienschwerpunkte : Wahrscheinlichkeitstheorie, math. Statistik, Masstheorie, Funktionalanalysis . Nebenfach : Physik
SS 76 - SS 79	: Studentischer Tutor am Fachbereich Mathematik der FUB
Juli 1979	: Diplompruefung in Mathematik . Thema der Diplomarbeit : Robuste Planung und Auswertung von Regressionsexperimenten
WS 79/80	: Tutor beim Projekt " Fernstudium im Medienverbund " am Fachbereich Mathematik der FUB
April 80 - Mai 84	: Wissenschaftlicher Mitarbeiter am Institut fuer angewandte Statistik am Fachbereich Wirtschaftswissenschaft der FUB . Taetigkeitsgebiet : Entwicklung von adaptiven Bayes'schen Stichprobensystem im Rahmen des DFG-gefoerderten Forschungsprojektschwerpunkts " Kostenoptimale adaptive Stichprobensysteme "